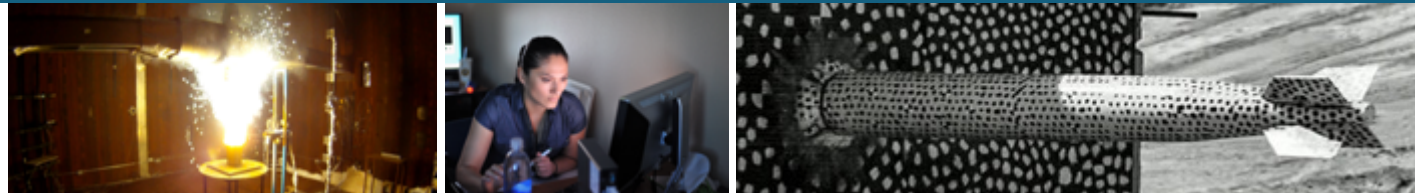




Application of Multifidelity Uncertainty Quantification Methods to a Subsurface Transport Model



SIAM CSE21

T. Portone, M.S. Eldred, G. Geraci, and L.P. Swiler

This work has been funded by the Geologic Disposal Safety Assessment (GDSA) program, under the Spent Fuel and Waste Science and Technology Campaign of the Office of Nuclear Energy in the U.S. Dept. of Energy.



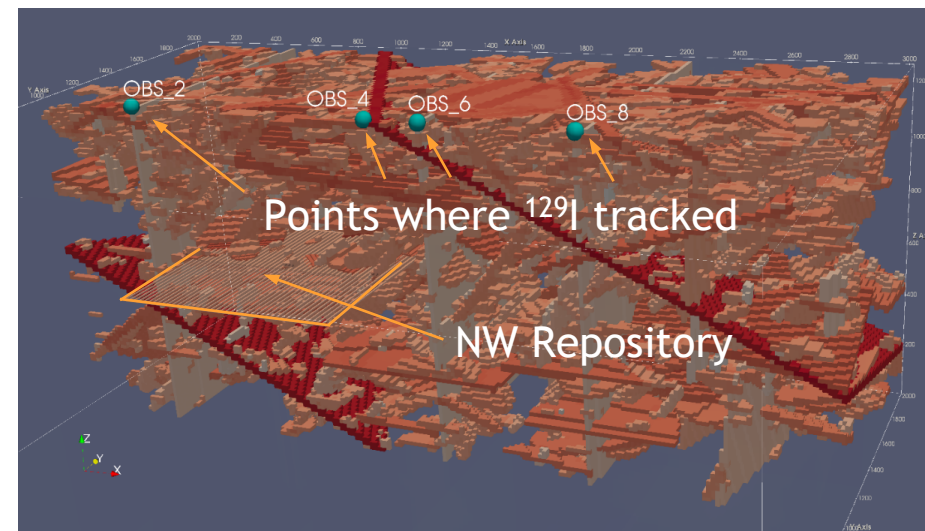
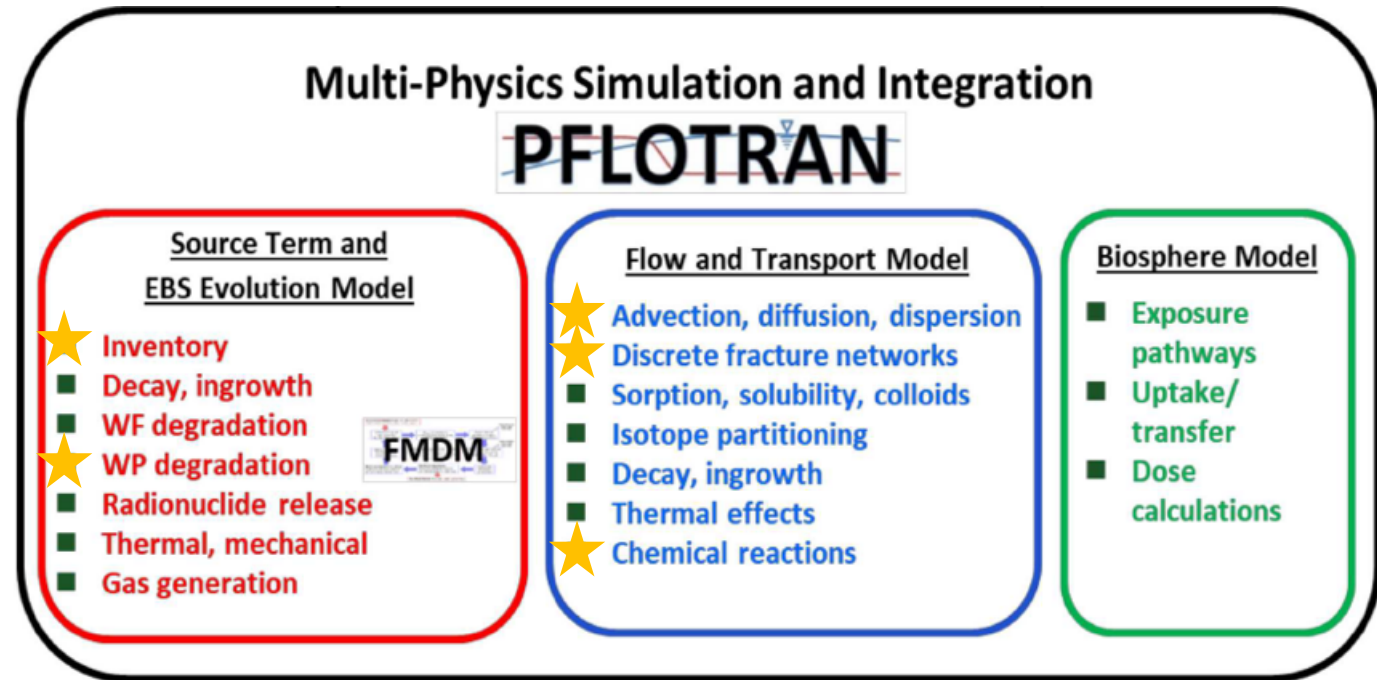
Sandia National Laboratories is a multission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

2 UQ for geologic disposal safety assessment (GD SA)



Performance assessment for nuclear waste repository site exhibits many computational challenges.

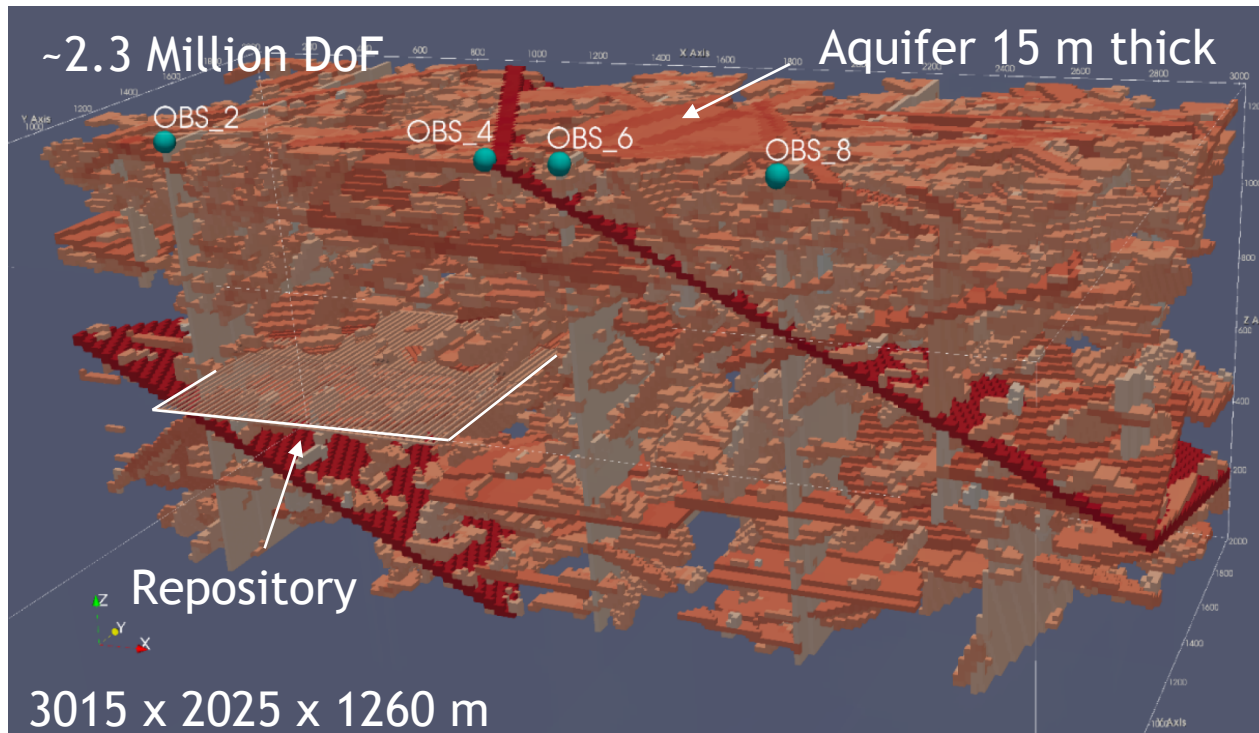
- Multiscale and multiphysics. Rife with uncertainties.
- Want to understand things like:
 - 1) subsurface/waste storage properties most important to repository performance;
 - 2) Probability dangerous levels of leaked waste would make it to a water supply.
- Deploying an unprecedented level of model fidelity for UQ studies in this application area.
- Simulations very costly: $\mathcal{O}(1000)$ model evaluations for production UQ studies.
- **Goal:** explore feasibility of using multifidelity polynomial chaos expansion (PCEs) for global sensitivity analysis (GSA) in this problem.



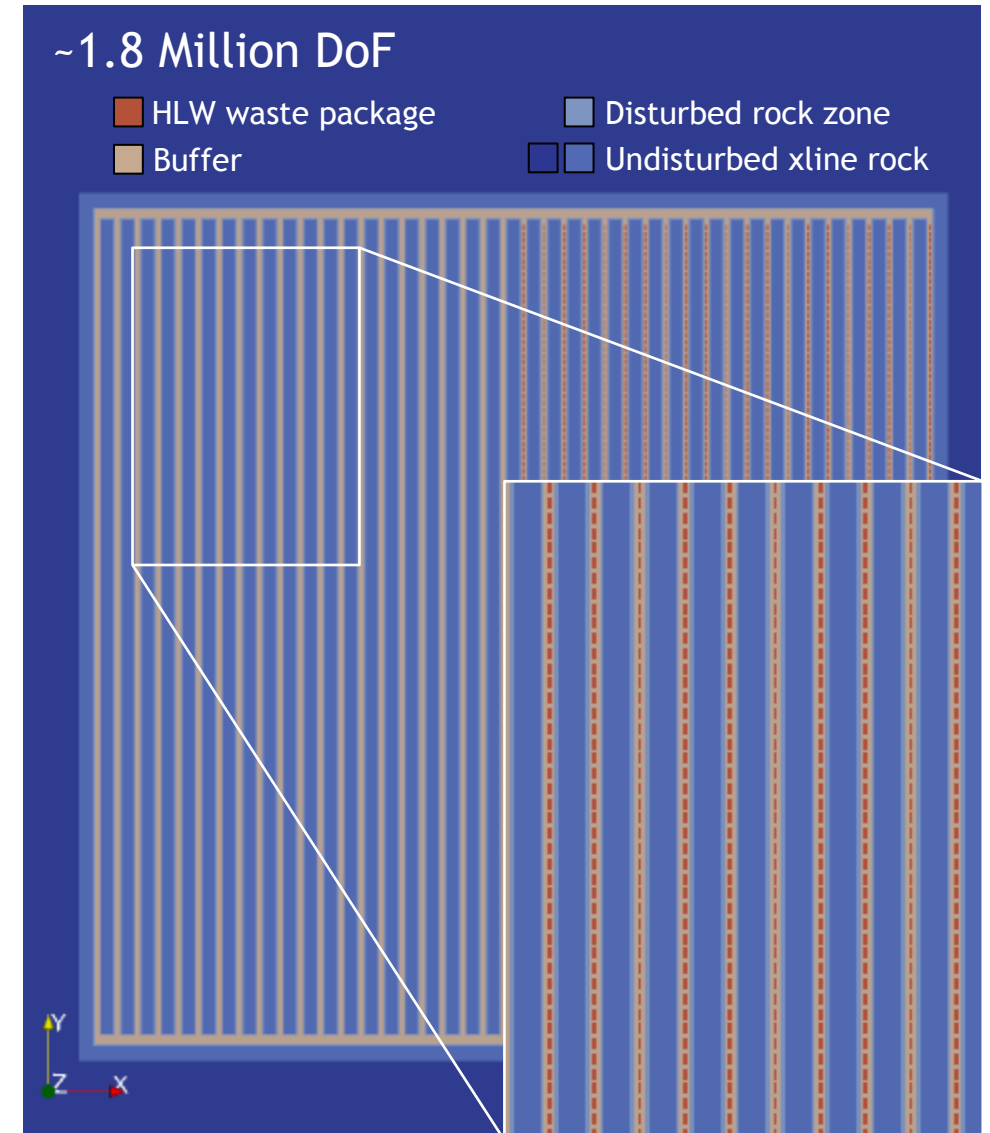
Production study crystalline reference case



- Considering uncertainties in:
 - Structure of subsurface (fracture locations/orientations/sizes)
 - Permeabilities/porosities of buffer, disturbed rock zone, aquifer
 - Waste package rates of degradation
- Simulation time 1 million years; 15 m uniform grid
- ~1.5 hours to run on 1152 cores



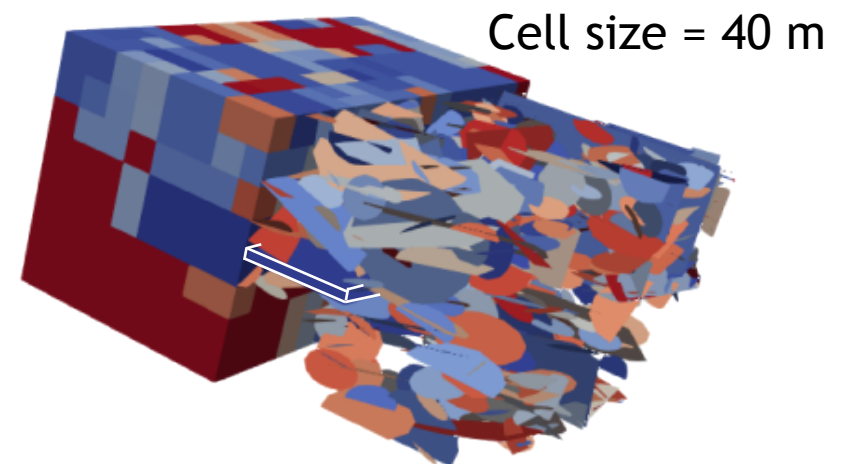
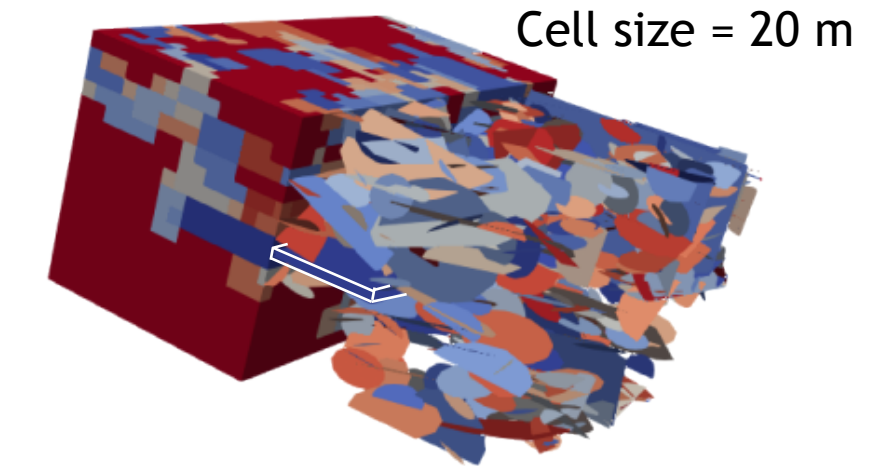
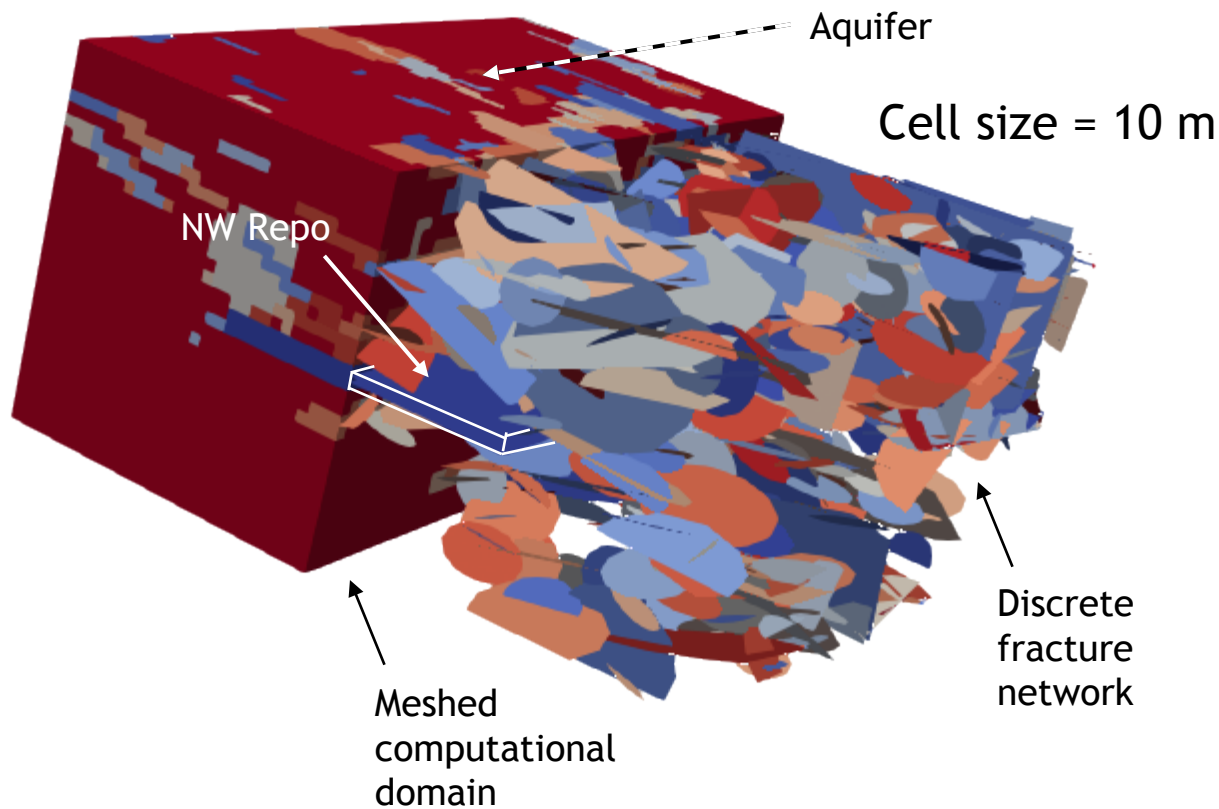
Meshed repository



Subsurface uncertainty treatment and model ensemble

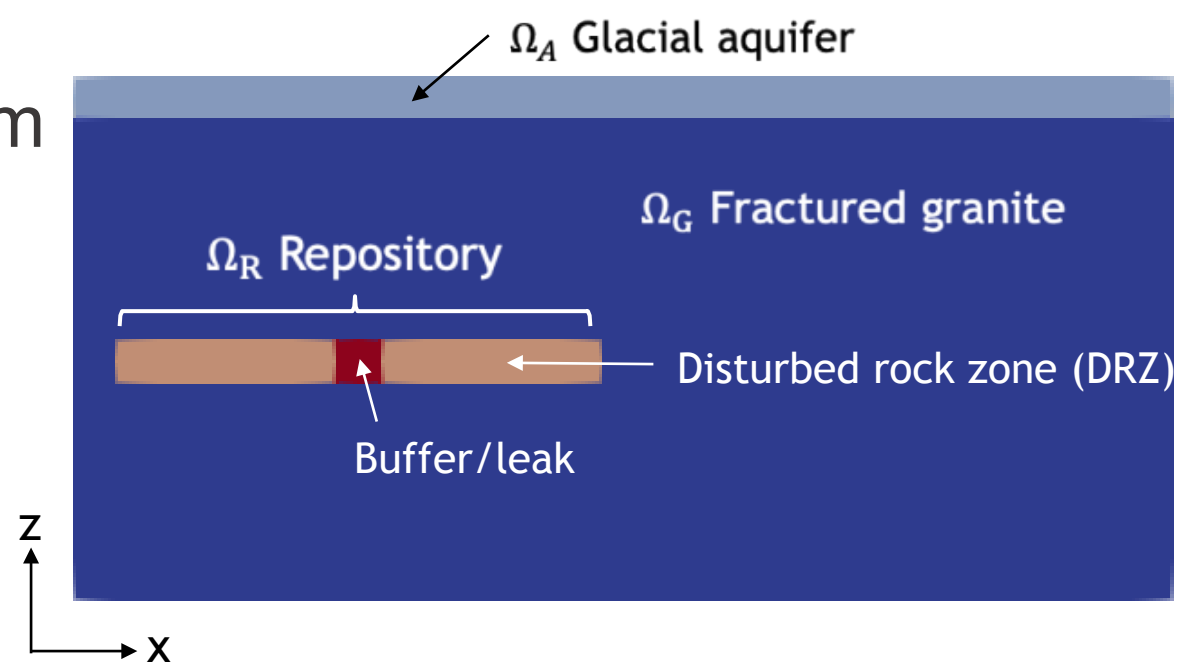


- Uncertainty in subsurface structure represented using random instantiations of discrete fracture networks (DFNs).
- DFNs mapped to equivalent continuous porous media (ECPMs).
 - Define regular grid, sweep through cells and map fracture permeability and porosity to equivalent continuum values.

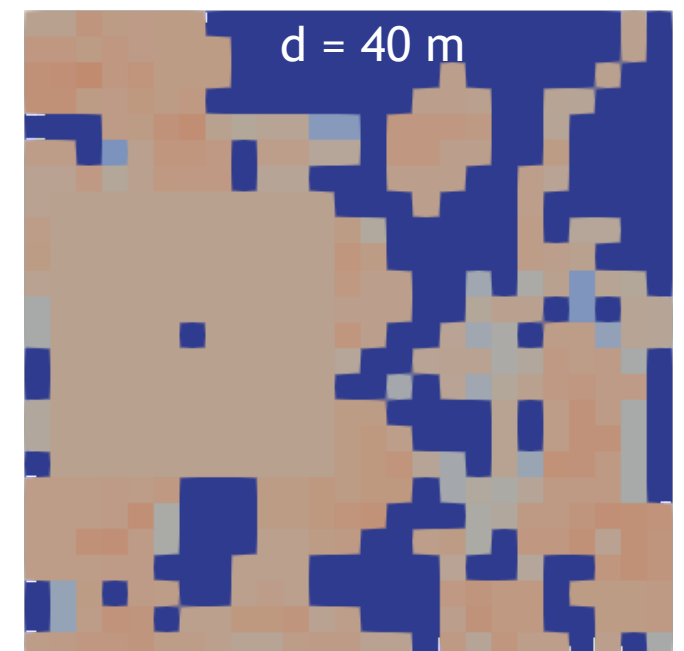
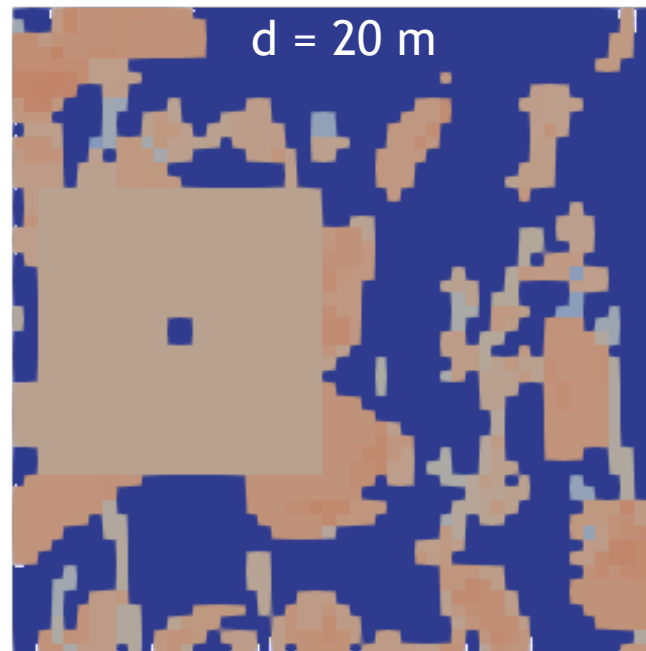
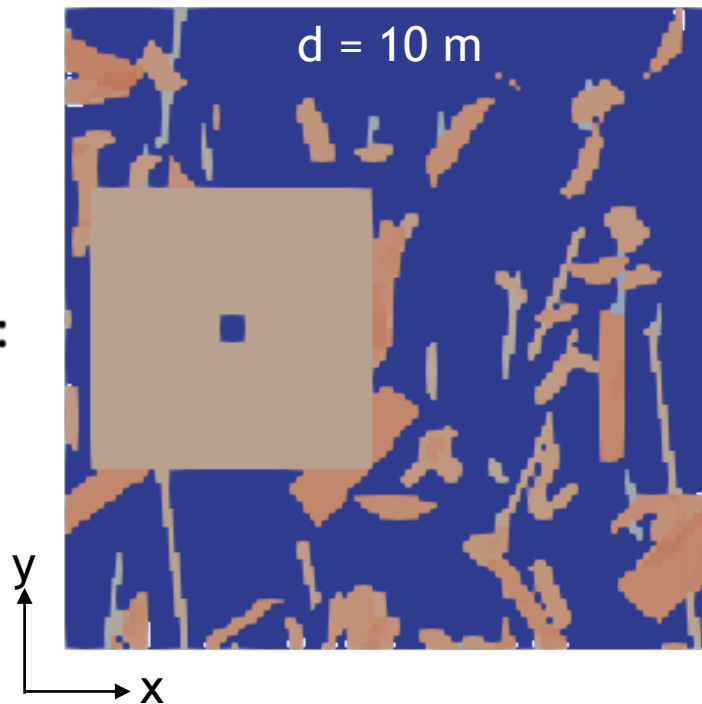


Simplified crystalline problem

- 1000 x 1000 x 500 m
- Run to 100 kyr rather than 1 Myr
- Drastically simplified repository structure
 - 1 waste package in buffer region
 - Not meshed separately
- Same underlying DFN statistics as production problem; min fracture radius 30 m.



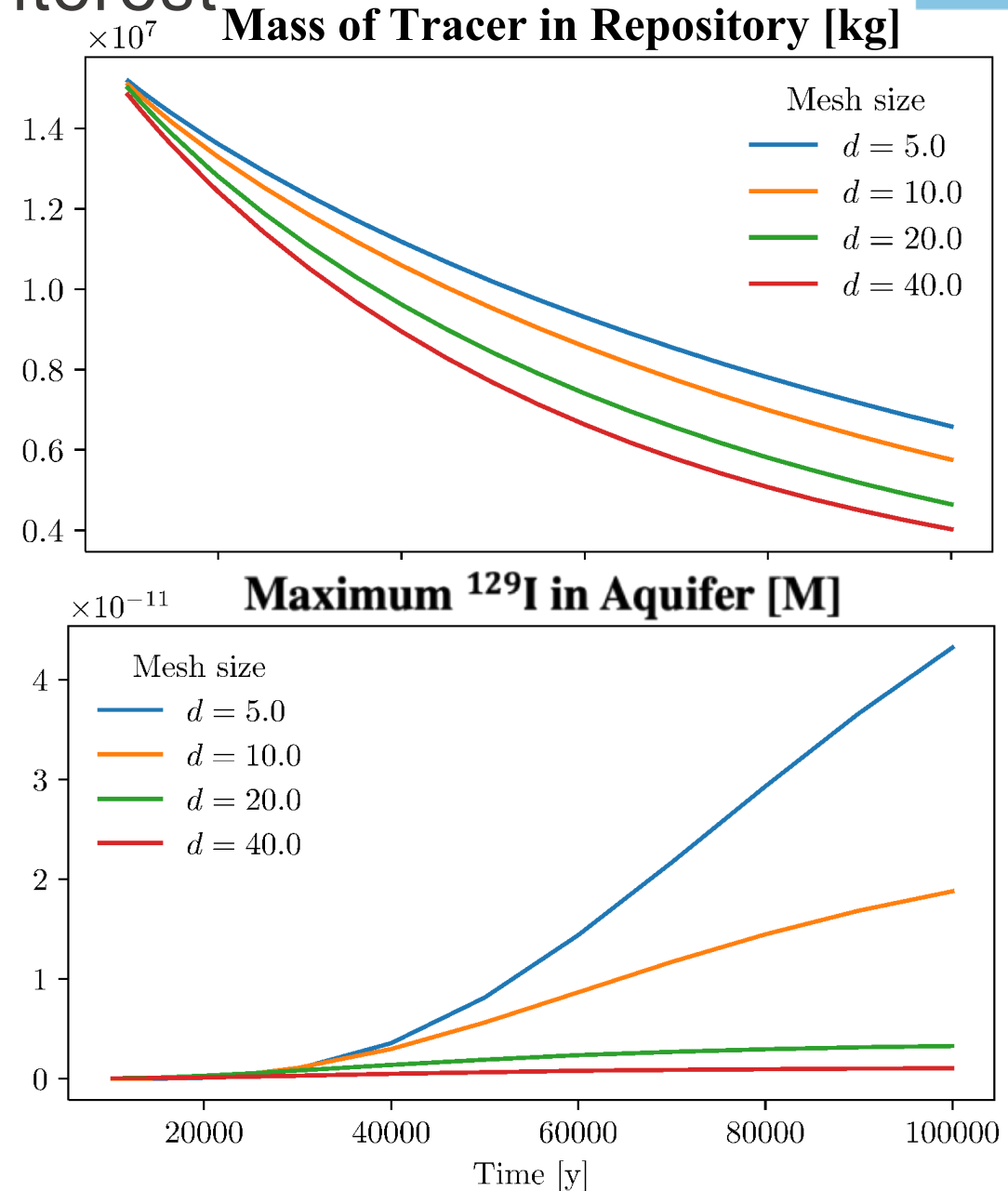
K_{xx} view
from above:



Mesh dependence of quantities of interest



- Some quantities of interest do not converge as mesh is refined
 - Tracking max ^{129}I concentration in aquifer and its location as a function of time
 - Location of max used for other derived quantities can vary significantly for coarser mesh (numerical diffusion/dispersion)
 - Tracking mass in repository; false connections from coarse mesh increase outflow
- Continuum values of ECPM depend on cell size
- Resolving minimum lengthscale (fracture aperture $\mathcal{O}(10^{-5} \text{ m})$) far beyond computational feasibility
- Using the finest mesh we can afford for UQ (15 m for production study)



Multifidelity Polynomial Chaos Expansion (PCE) for GSA



- Considering performance of discrepancy-based, hierarchical multifidelity PCE (Ng, Eldred 2012) as surrogate for high-fidelity model in GSA.
- Constructed using recursive discrepancies (R – model response; P – PCE expansion):

$$R_{10} = R_{10} + (R_{20} - R_{20}) + (R_{40} - R_{40}) = R_{40} + (R_{20} - R_{40}) + (R_{10} - R_{20})$$

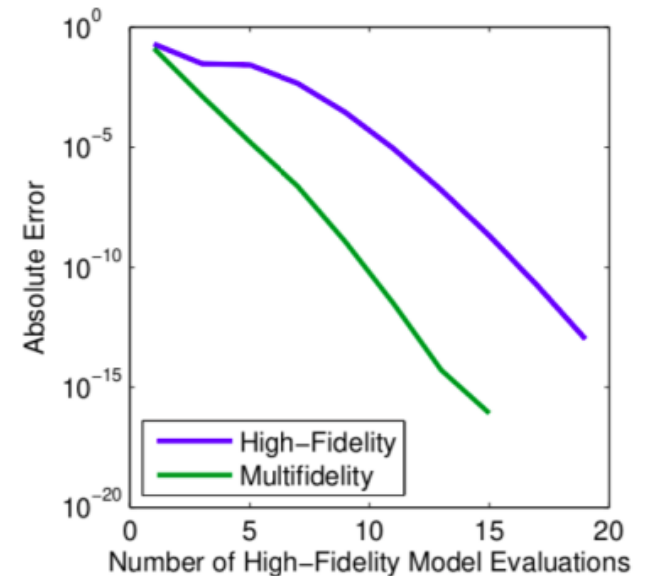
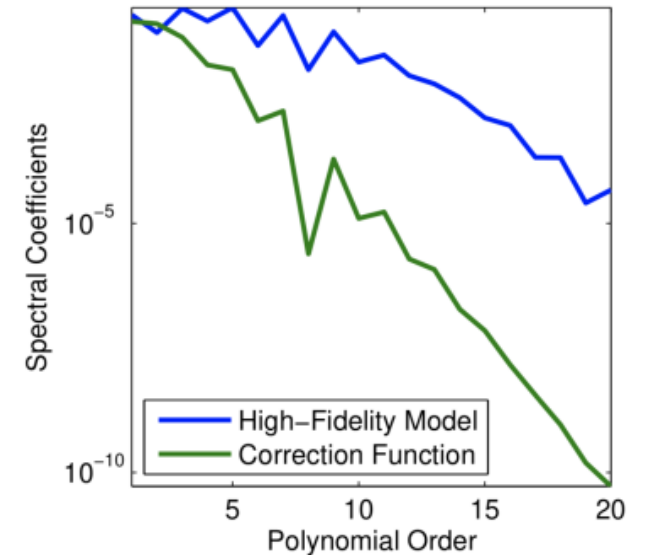
$$\approx P_{40} + (P_{20-40} + P_{10-20})$$

where $P_{20-40} \approx R_{20} - R_{40}$, etc.

- If spectral content of $P_{10-20} \ll R_{10}$, need fewer $d=10$ evaluations to derive accurate expansion of R_{10} (see example on right).
- In this case using a recursive approach to construct discrepancy PCE:

$$P_{20-40} \approx R_{20} - P_{40} \text{ rather than } P_{20-40} \approx R_{20} - R_{40}.$$

- Coefficients in each PCE can be recombined into a single PCE expansion for $d = 10$ and used to compute Sobol indices in postprocessing (Sudret 2008).
- Random (MC) sampling + coefficients estimated with regression + cross validation for PCE order.
- Sampling and PCE construction performed with Dakota.

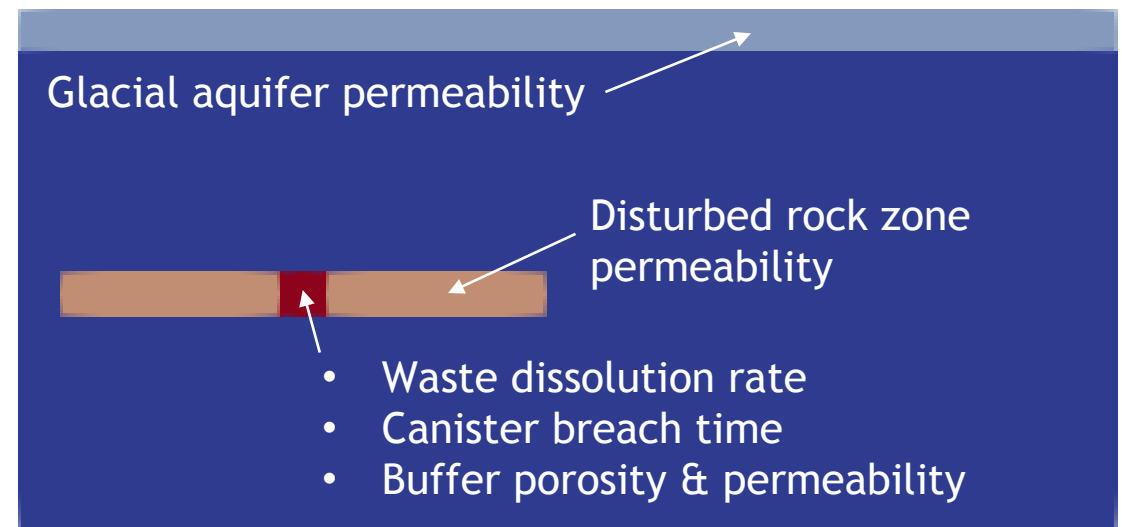


Preliminary multifidelity PCE (MF PCE) study, 1 DFN



- Used 1 DFN for all samples
 - Study variability caused by meshing without conflating variability from uncertain subsurface (size, location, density of fractures).
- Hand-selected number of samples at $d = 10, 20,$ and 40 for preliminary studies.
- Generated reference solution with 828 samples on finest mesh.
- Interested in potential benefits of augmenting small number of high-fidelity (HF) samples with samples from coarser meshes.
- Generated single-fidelity PCEs with varying numbers of HF samples for comparison to multifidelity PCE.
- Generated multifidelity PCEs augmenting HF samples with 84 samples for $d = 20, d = 40$.

	$d = 10$	$d = 20$	$d = 40$
N samples	18	84	84
Relative costs	1	0.02	0.006

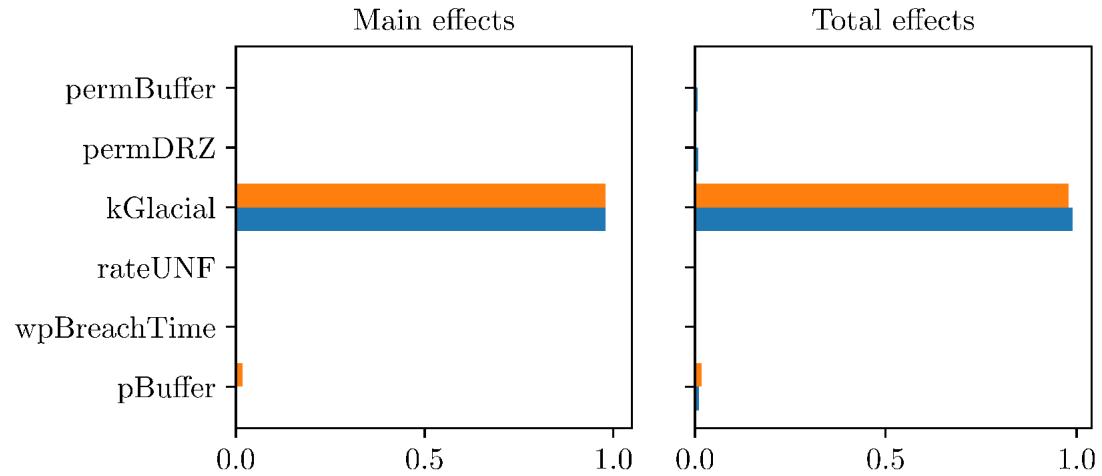


Sobol indices for QOI with simple sensitivities are well resolved with only 18 HF model evaluations.

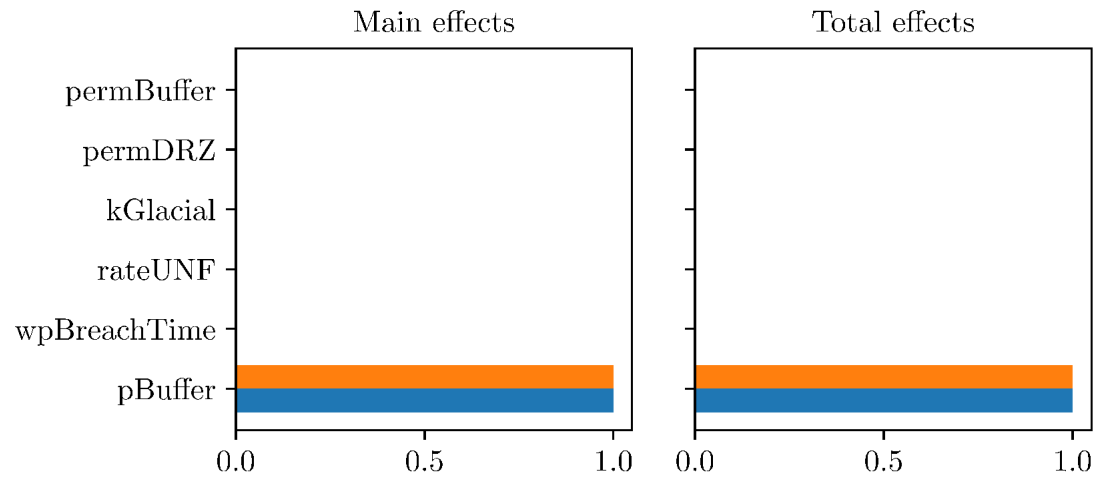


Single-fidelity PCE, HF model

$\max_{\underline{x} \in \Omega_A} {}^{129}\text{I}$ [M] at 100 kyr



\max_t Mean Residence Time of Repository [yr]

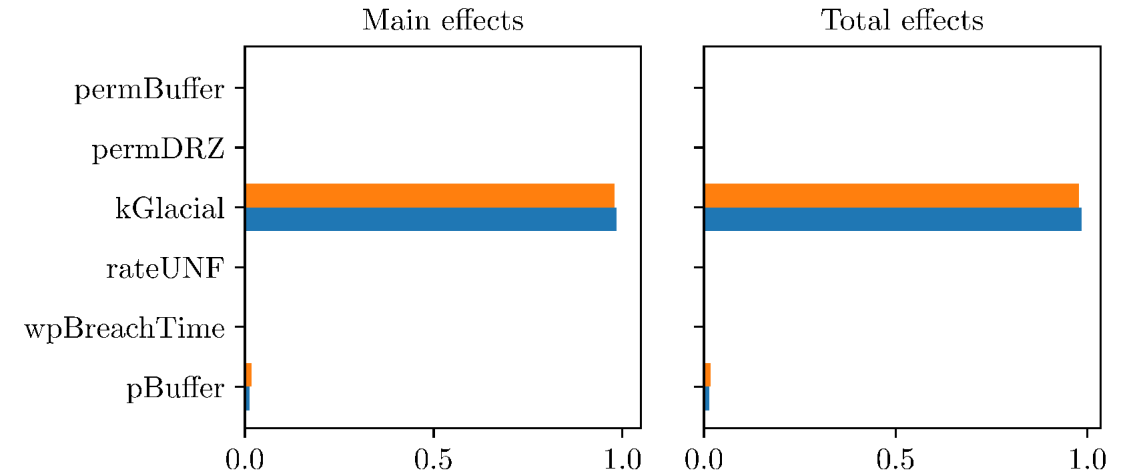


N Equivalent HF samples

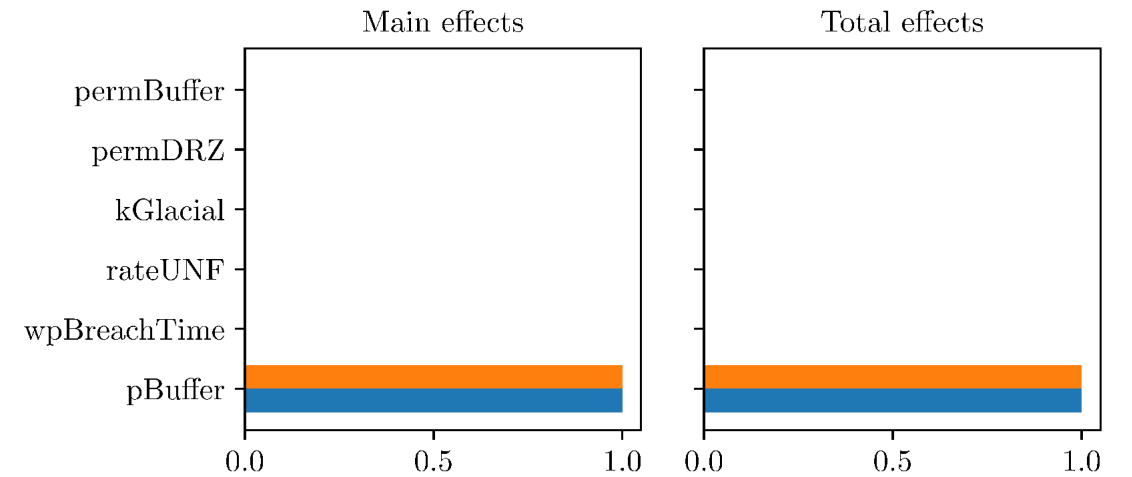


Multifidelity PCE

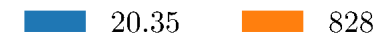
$\max_{\underline{x} \in \Omega_A} {}^{129}\text{I}$ [M] at 100 kyr



\max_t Mean Residence Time of Repository [yr]



N Equivalent HF samples

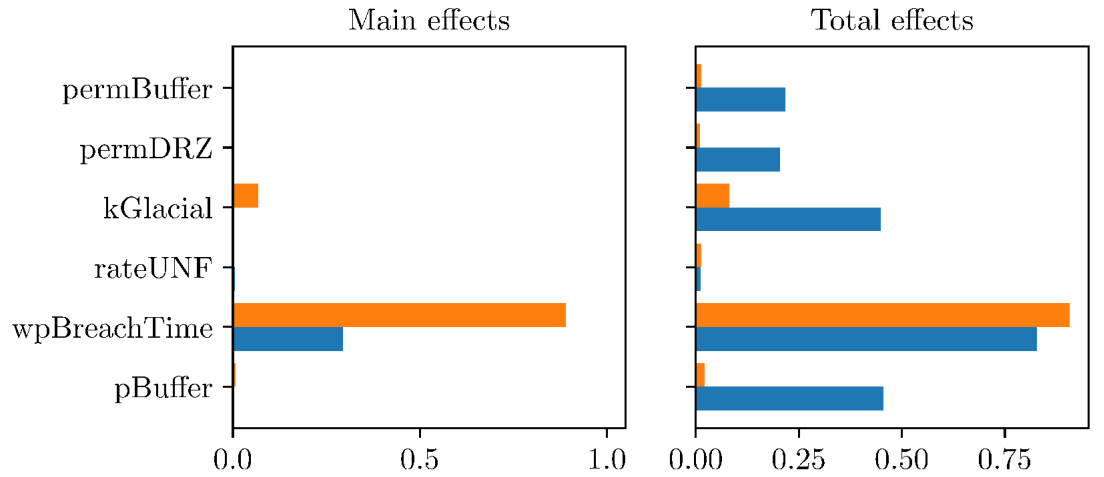


Some Sobol indices with more complex dependencies were improved by employing MF PCE; however, some (bottom) were not, because of large differences in behavior based on mesh.

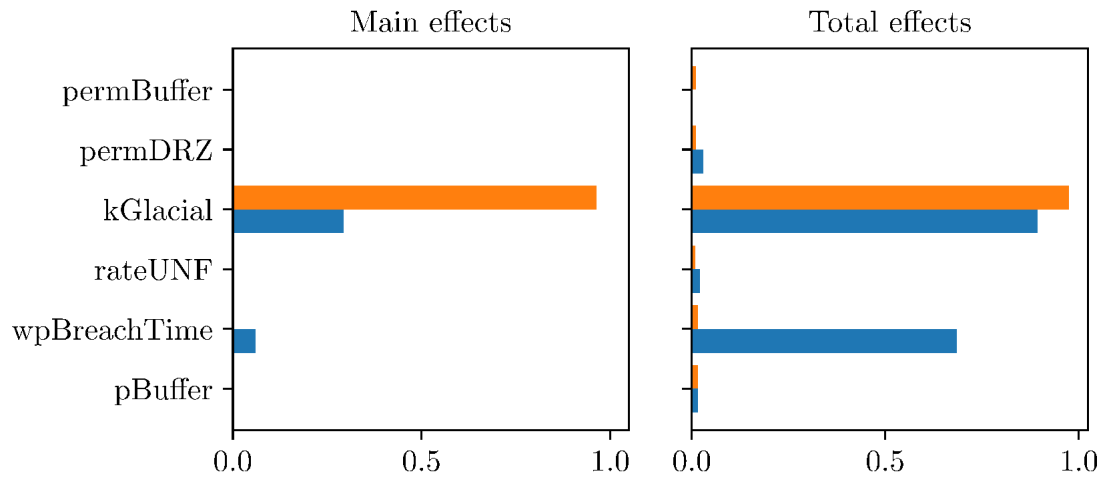


Single-fidelity PCE, HF model

$\max_{\mathbf{x} \in \Omega_A} {}^{129}\text{I}$ [M] at 10 kyr



x location of Peak ${}^{129}\text{I}$ [m]

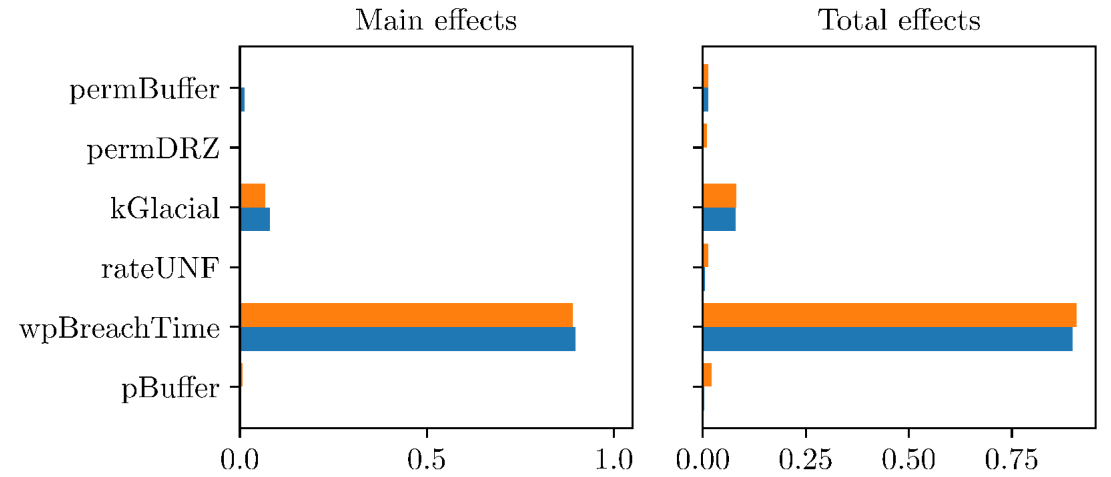


N Equivalent HF samples

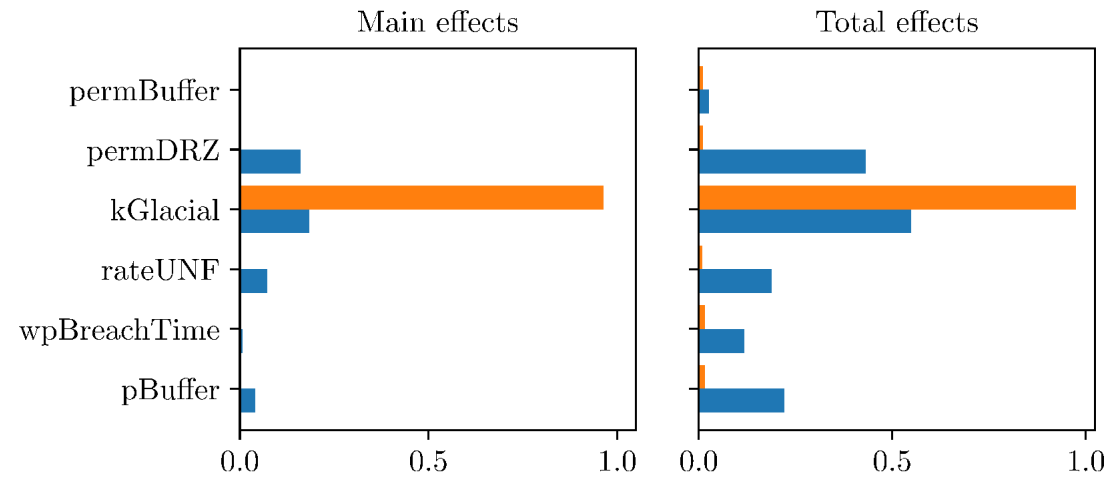
18 828

Multifidelity PCE

$\max_{\mathbf{x} \in \Omega_A} {}^{129}\text{I}$ [M] at 10 kyr



x location of Peak ${}^{129}\text{I}$ [m]



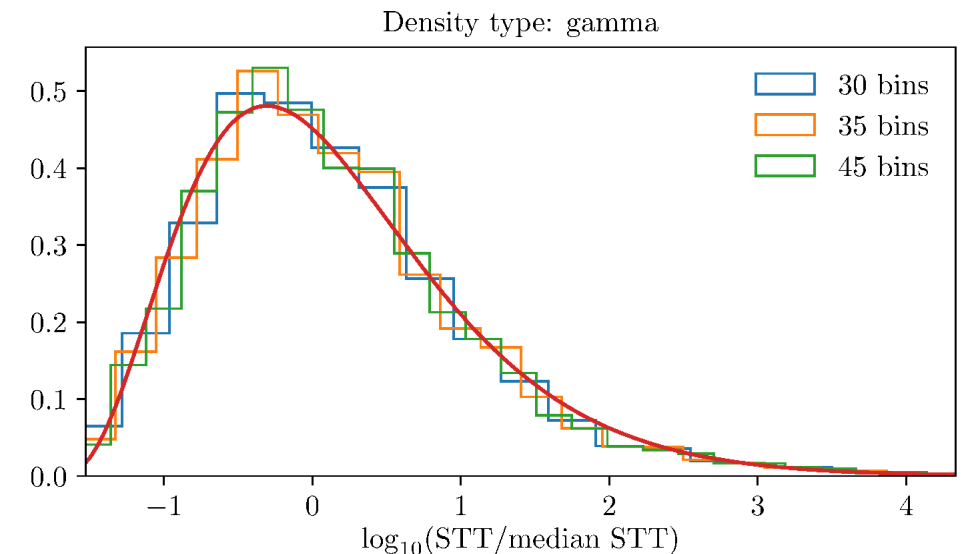
N Equivalent HF samples

20.35 828

Characterizing subsurface uncertainty



- How do properties of fractured granite affect quantities of interest through GSA?
- Compute proxy variables based on graph/network representation of each DFN (fractures = nodes, intersections = edges)
 - Number intersections with repository (NIwR)
 - Average degree (average number of intersections a fracture is part of)
 - Shortest travel time (STT) between repo and aquifer (rough approx. based on fracture area and pressure differentials)
- These are derived quantities and rough approximations of underlying subsurface properties.
 - r^2 values ≥ 0.20 between these proxy variables and several QOIs
 - They are derived and therefore not controllable by many adaptive algorithms; use regression
 - Augment build point input space with computed values of proxy variables for each DFN
 - Derived empirical distributions to define PCE bases from ~ 3000 samples of proxy variables
- **This is the first time we have been able to account for subsurface variability in GSA**
 - However, these are only proxies and cannot completely capture the induced variance from subsurface.



Preliminary multifidelity PCE study with subsurface uncertainty



- Each sample model evaluation was generated with a different DFN (different subsurface instantiation)
- Reference PCE solution workflow:
 - Generate 828 sample model evaluations on finest mesh with different DFN for each sample
 - Augment input variable samples with proxy variable values from each DFN
 - Construct regression PCE over augmented input space
- Multifidelity PCE workflow proceeds in similar manner
 - Augmented varying numbers of HF samples with 828 samples for $d = 20$, $d = 40$.

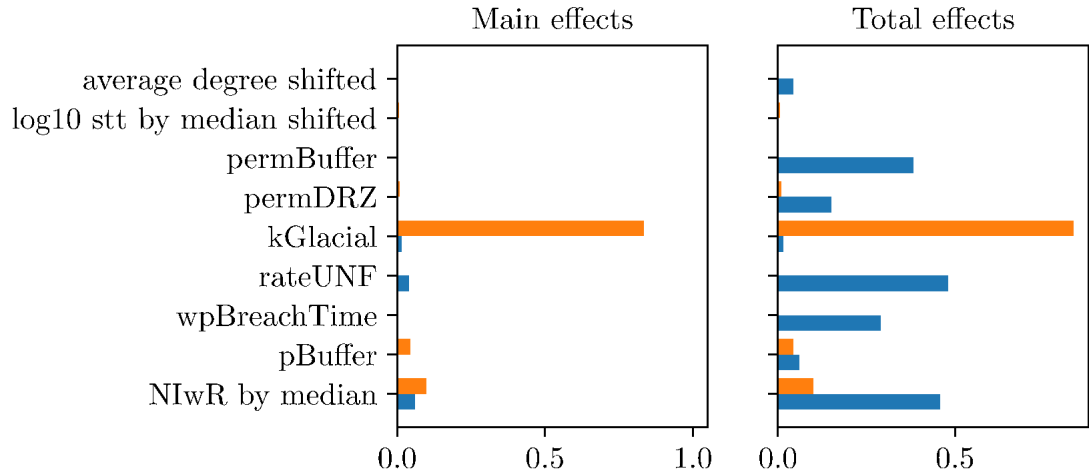
	d = 10	d = 20	d = 40
N samples	18, 576	828	828
Relative costs	1	0.02	0.006

For all but the QOIs depending on location of peak ^{129}I , estimation of Sobol indices was drastically improved with the multifidelity PCE.

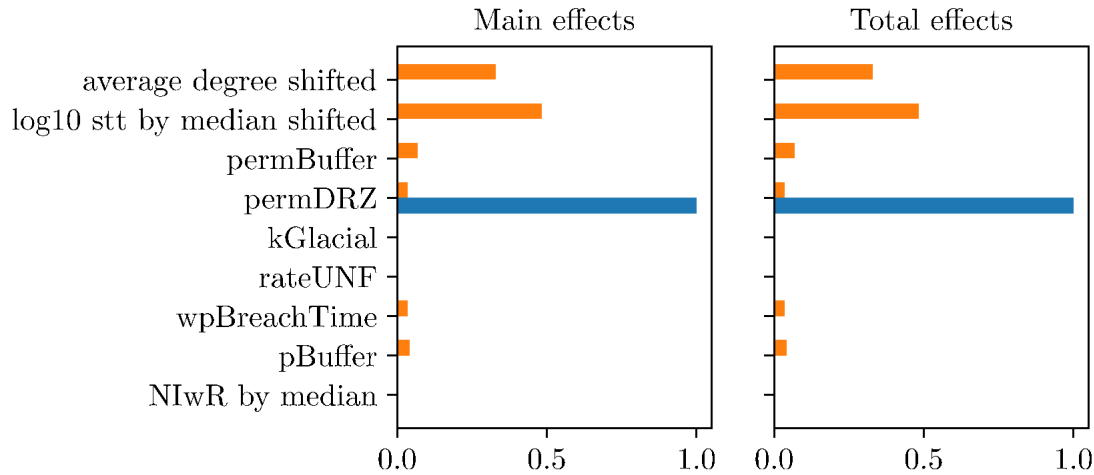


Single-fidelity PCE, HF model

$$\text{Peak } ^{129}\text{I} \equiv \max_{t, \mathbf{x} \in \Omega_A} ^{129}\text{I} [\text{M}]$$



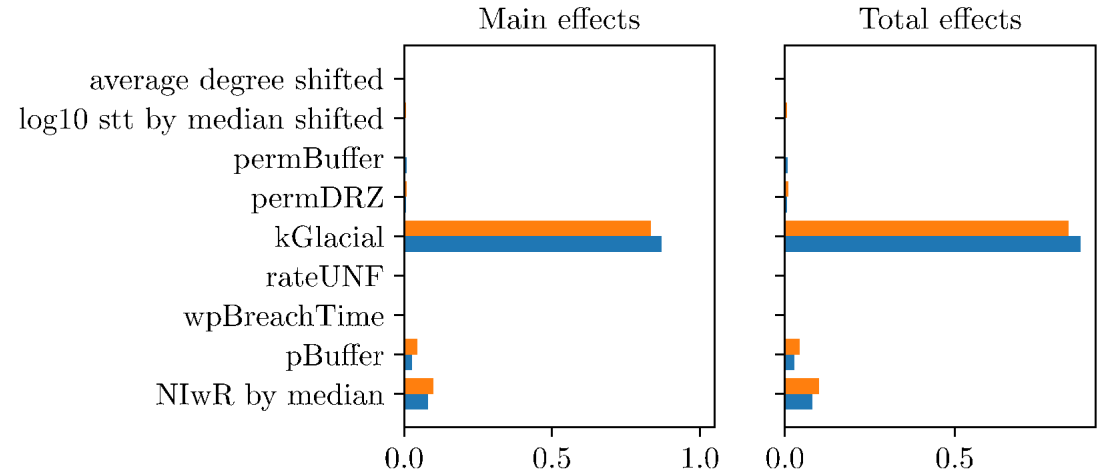
$$\max_t \text{ Mass Flux from Repository } [\text{kg/yr}]$$



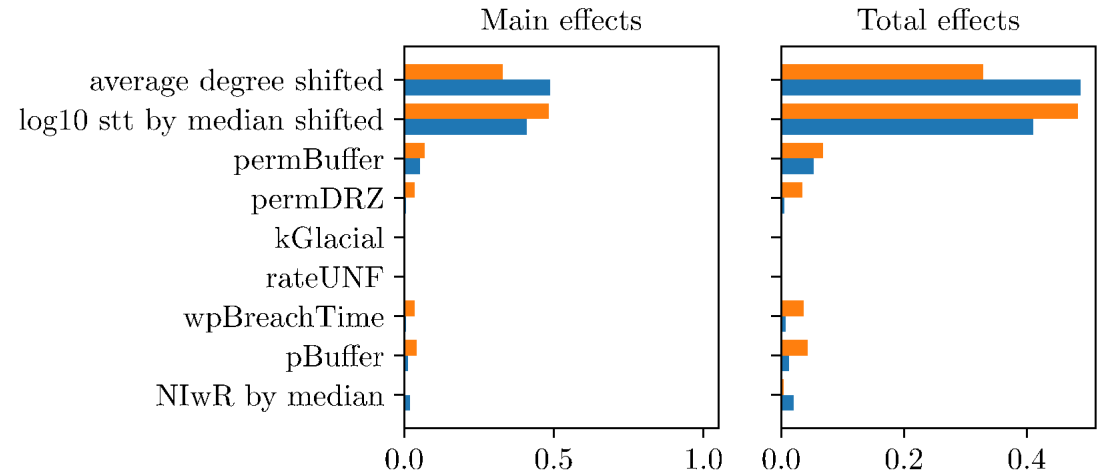
N Equivalent HF samples
■ 18 ■ 828

Multifidelity PCE

$$\text{Peak } ^{129}\text{I} \equiv \max_{t, \mathbf{x} \in \Omega_A} ^{129}\text{I} [\text{M}]$$



$$\max_t \text{ Mass Flux from Repository } [\text{kg/yr}]$$



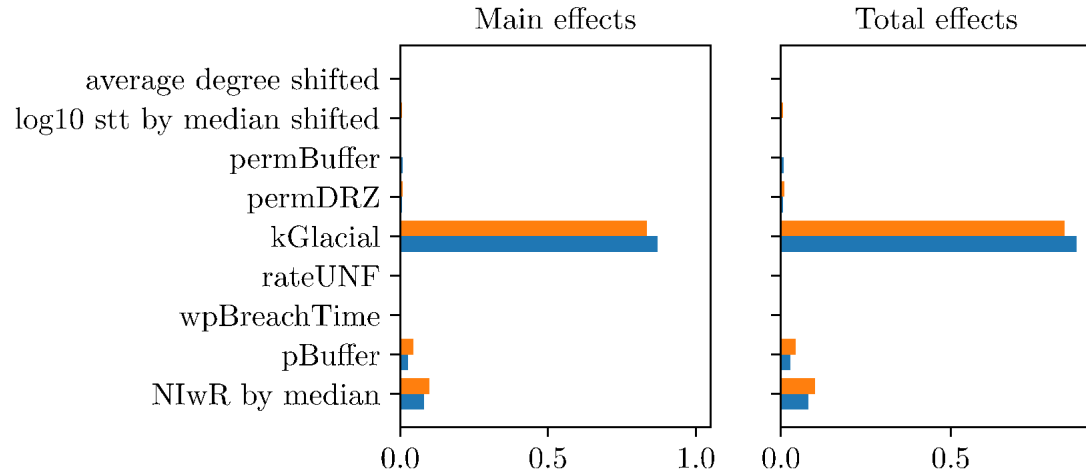
N Equivalent HF samples
■ 41.17 ■ 828

Precision of Sobol indices computed from the multifidelity PCE did not improve much with increase HF evaluations.

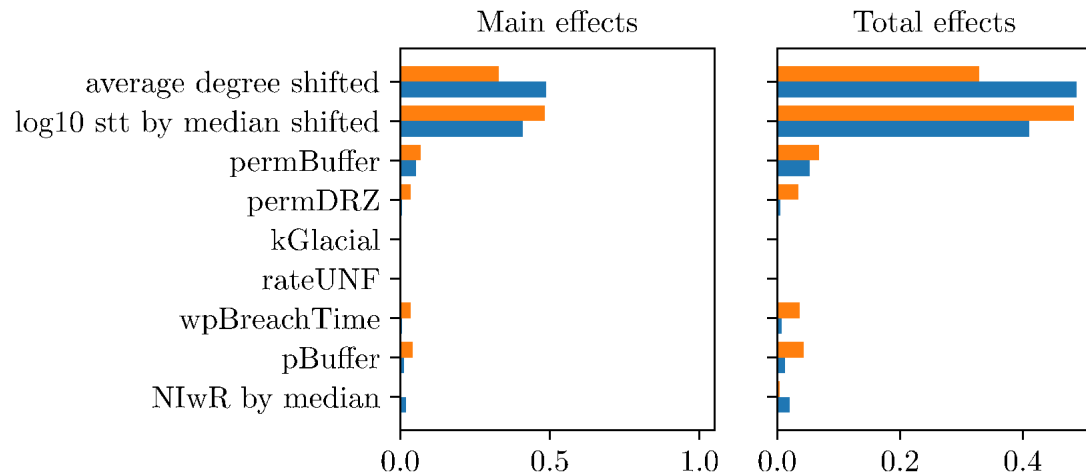


Single-fidelity PCE, HF model

Peak $^{129}\text{I} \equiv \max_{t, \mathbf{x} \in \Omega_A} ^{129}\text{I} [\text{M}]$



\max_t Mass Flux from Repository [kg/yr]

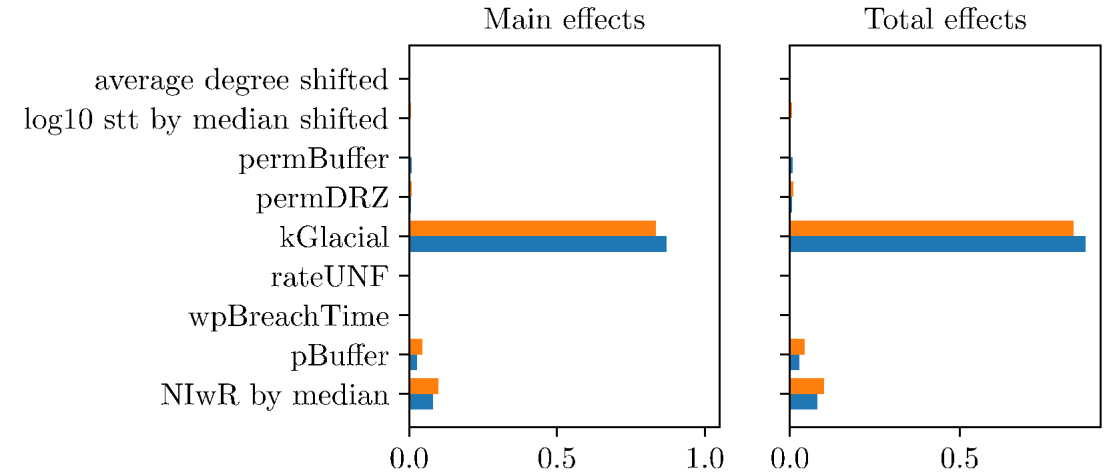


N Equivalent HF samples

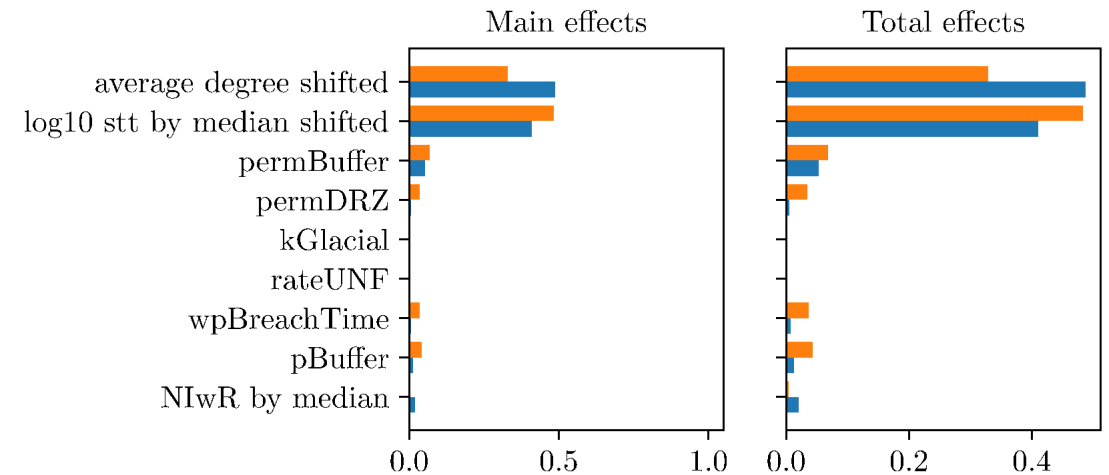
41.17 828

Multifidelity PCE

Peak $^{129}\text{I} \equiv \max_{t, \mathbf{x} \in \Omega_A} ^{129}\text{I} [\text{M}]$



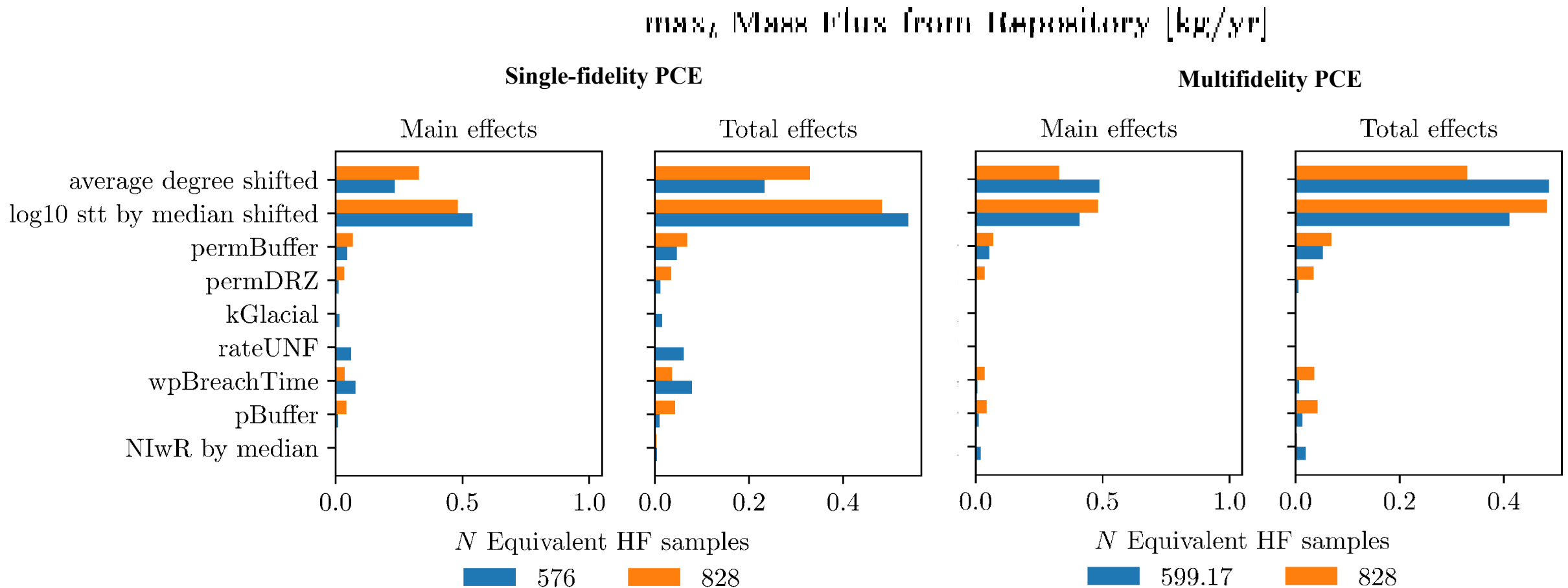
\max_t Mass Flux from Repository [kg/yr]



N Equivalent HF samples

599.17 828

- Sobol indices not significantly improved for MF PCE vs. single-fidelity PCE using high-fidelity only in some cases.
- Need to inspect PCE reconstructions at each level in more detail to determine why this is happening.



Takeaways/future directions



- Overall, these results show there is promise in pursuing MF PCE for GSA in this problem.
- However, performance for QoIs depending significantly on mesh was poor.
- Need to determine reasonable model hierarchy that is useful for all QoIs.
- Upscaling assumption that all fracture diameters are \geq cell size is violated for $d = 20, 40$.
 - Filter out smallest fractures?
- Use adaptive mesh refinement to mitigate false connections and improve resolution of fractures?
- Redefine model hierarchy in other forms without modifying subsurface mesh?
 - Maximum timestep allowed in global implicit timestepper used to advance the model.
 - In production run, coarsen repository mesh instead.
- Consider augmenting with few finer samples vs many coarser samples.
- Optimal sample allocations for approximating Sobol indices with multifidelity PCE – see Michael Merritt’s talk later in the session!