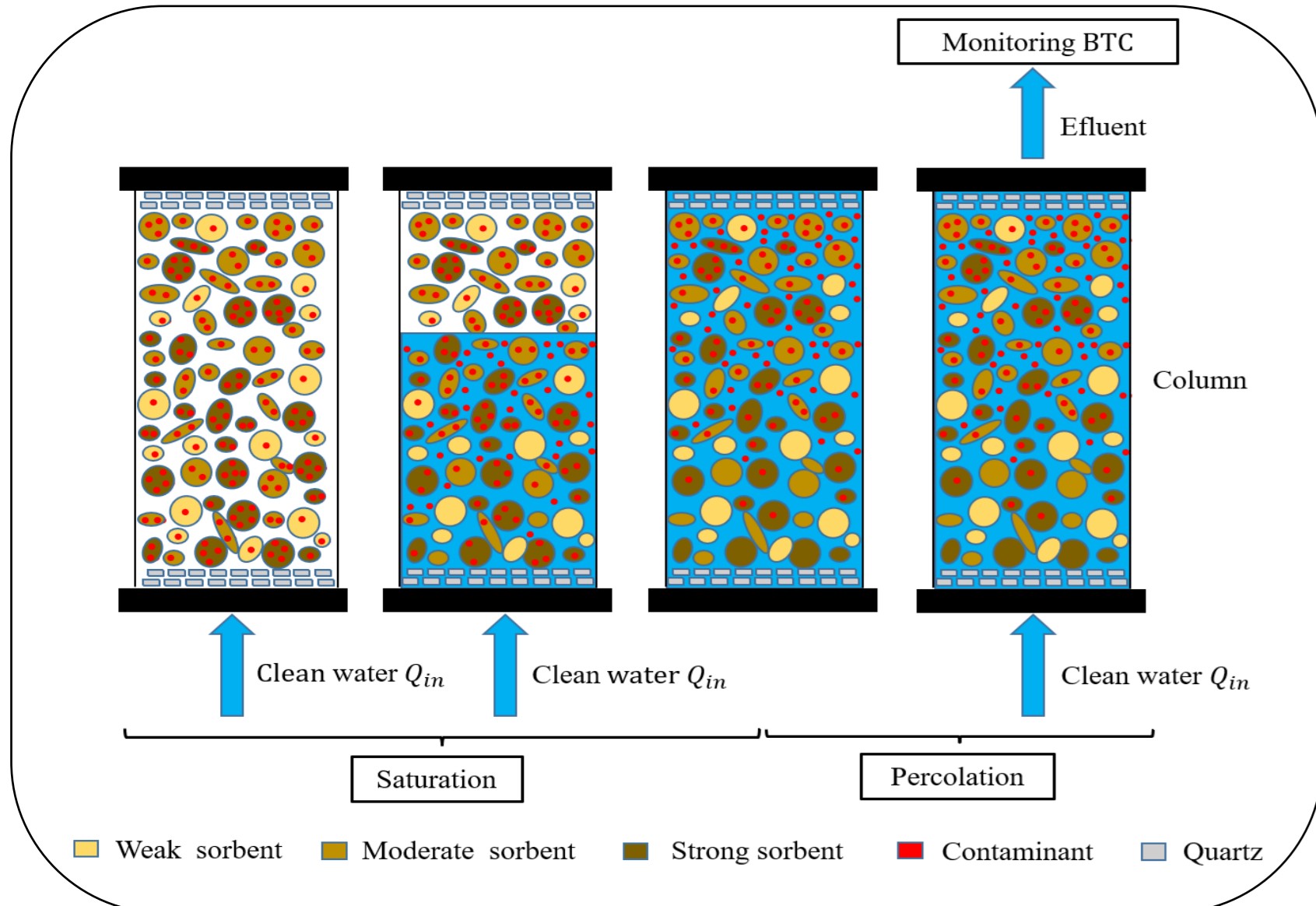
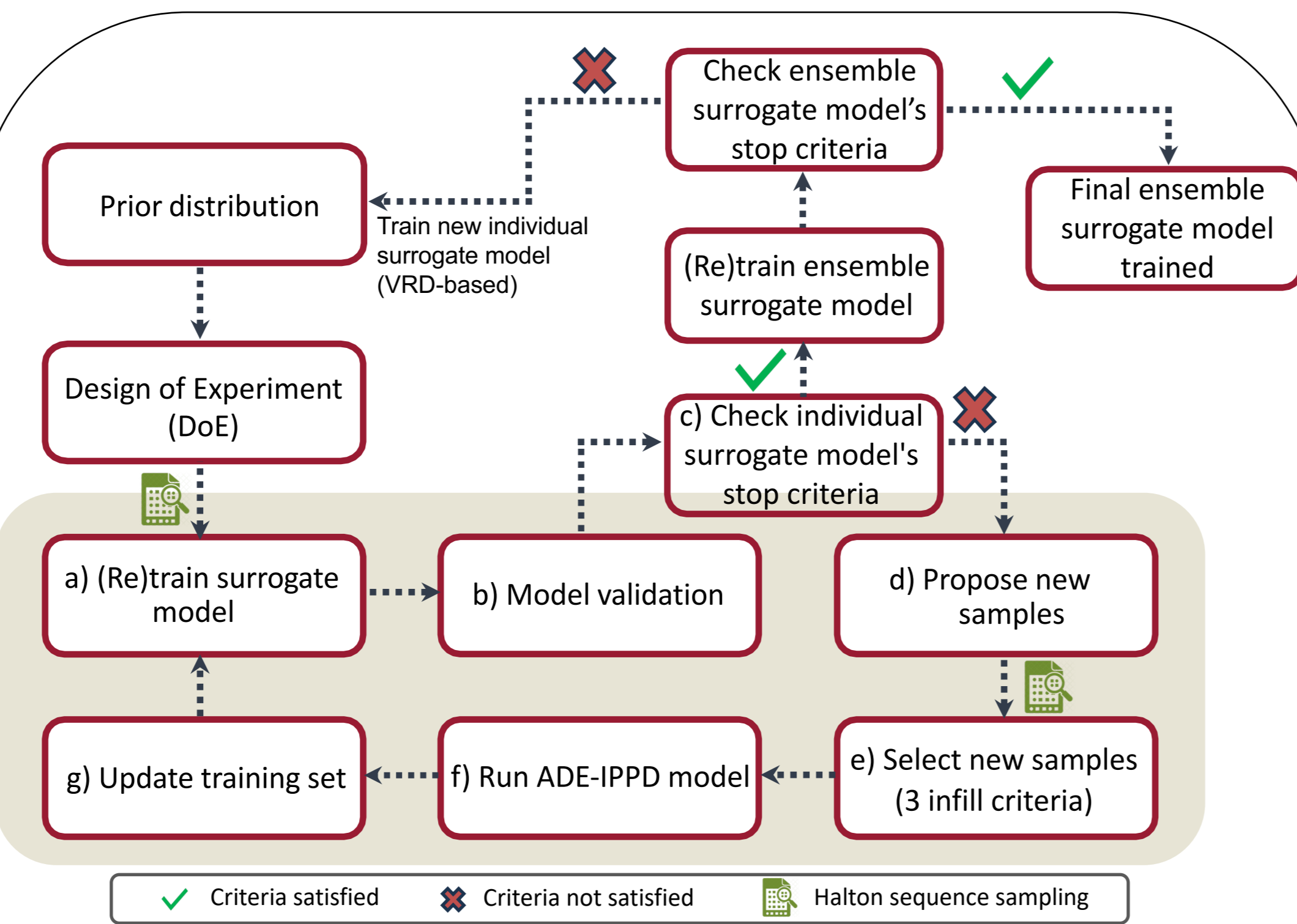


KEY IDEA

Setup of up-flow column leaching test

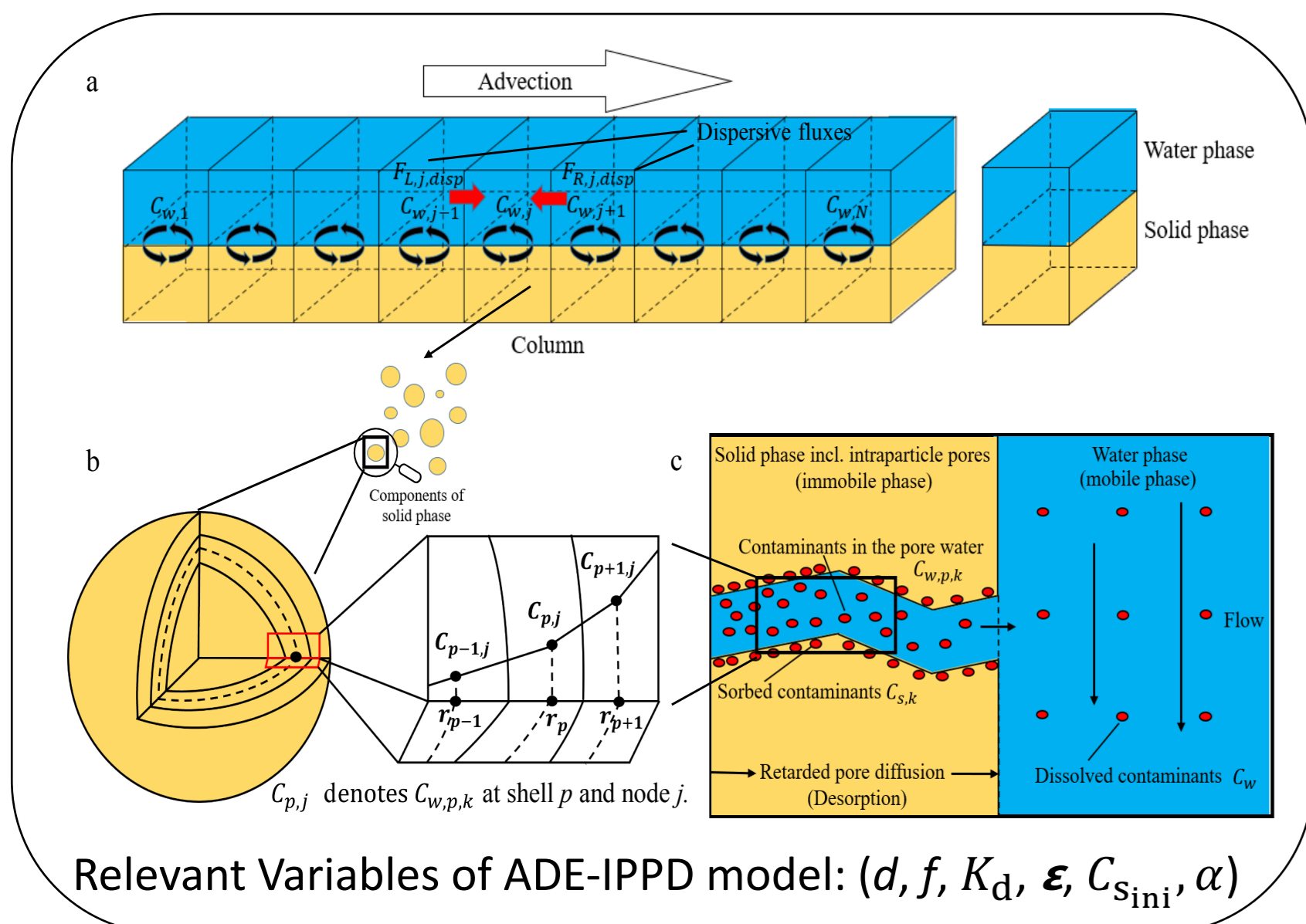


Creation of ensemble surrogate models



- Initial samples (DoE).
- Individual surrogate model stop: RRMSE < 0.1 (relevant VRD) or training set > 1100 samples.
- Ensemble surrogate model stop: RRMSE < 0.1 (test set) or all 200 surrogate models are trained.
- 3 infill criteria: Maximize standard deviation for exploration. Maximize Expected Improvement and minimize Mahalanobis distances for exploitation.

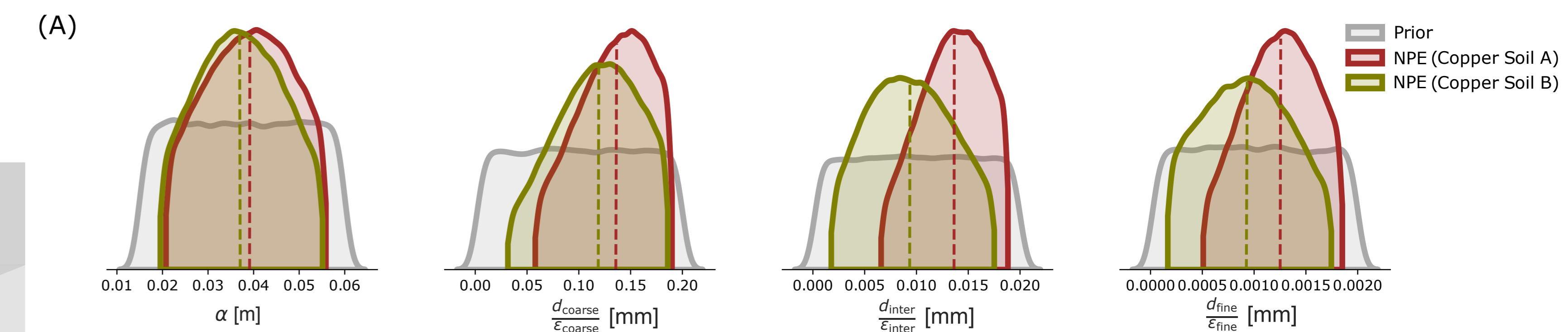
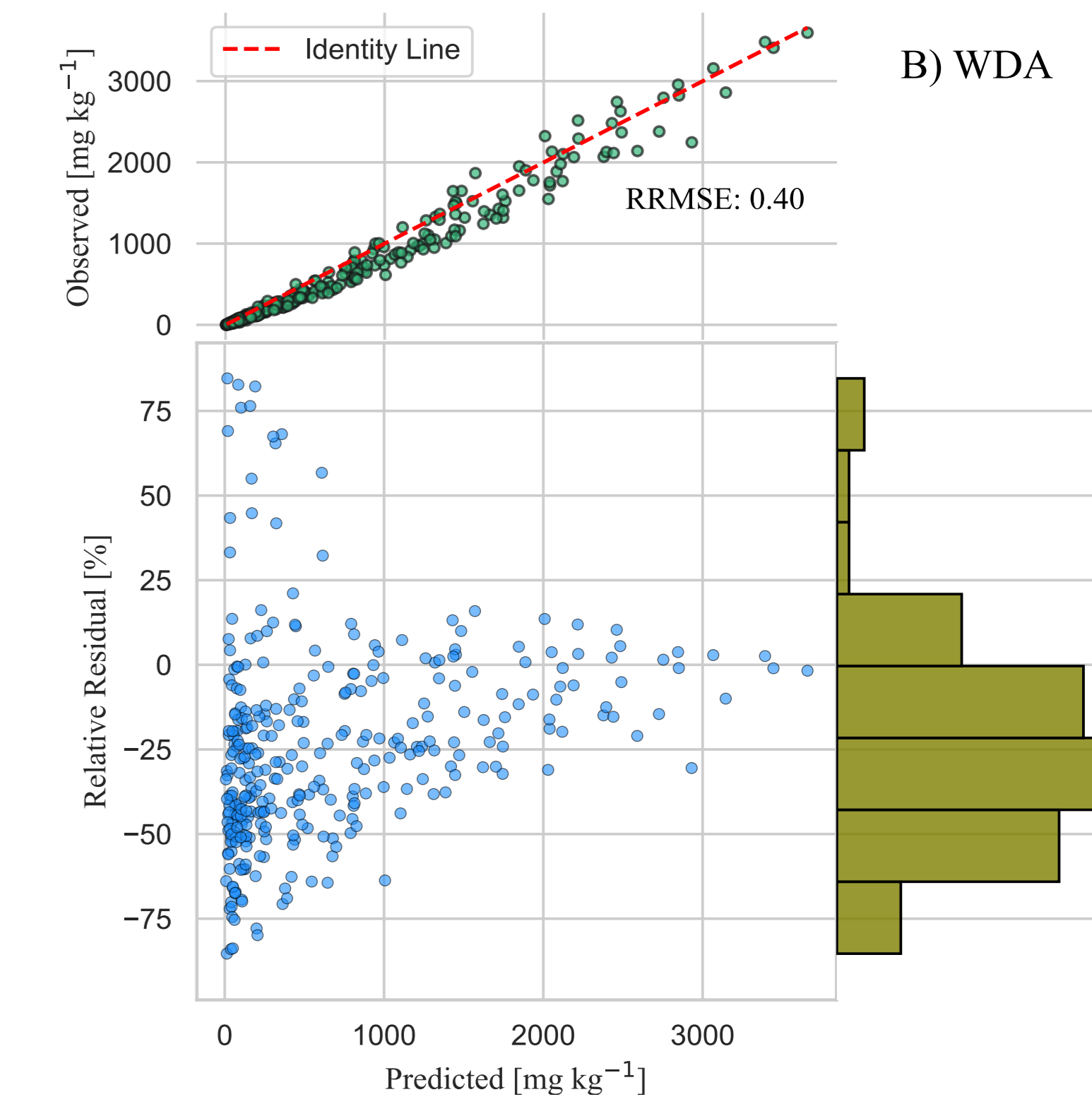
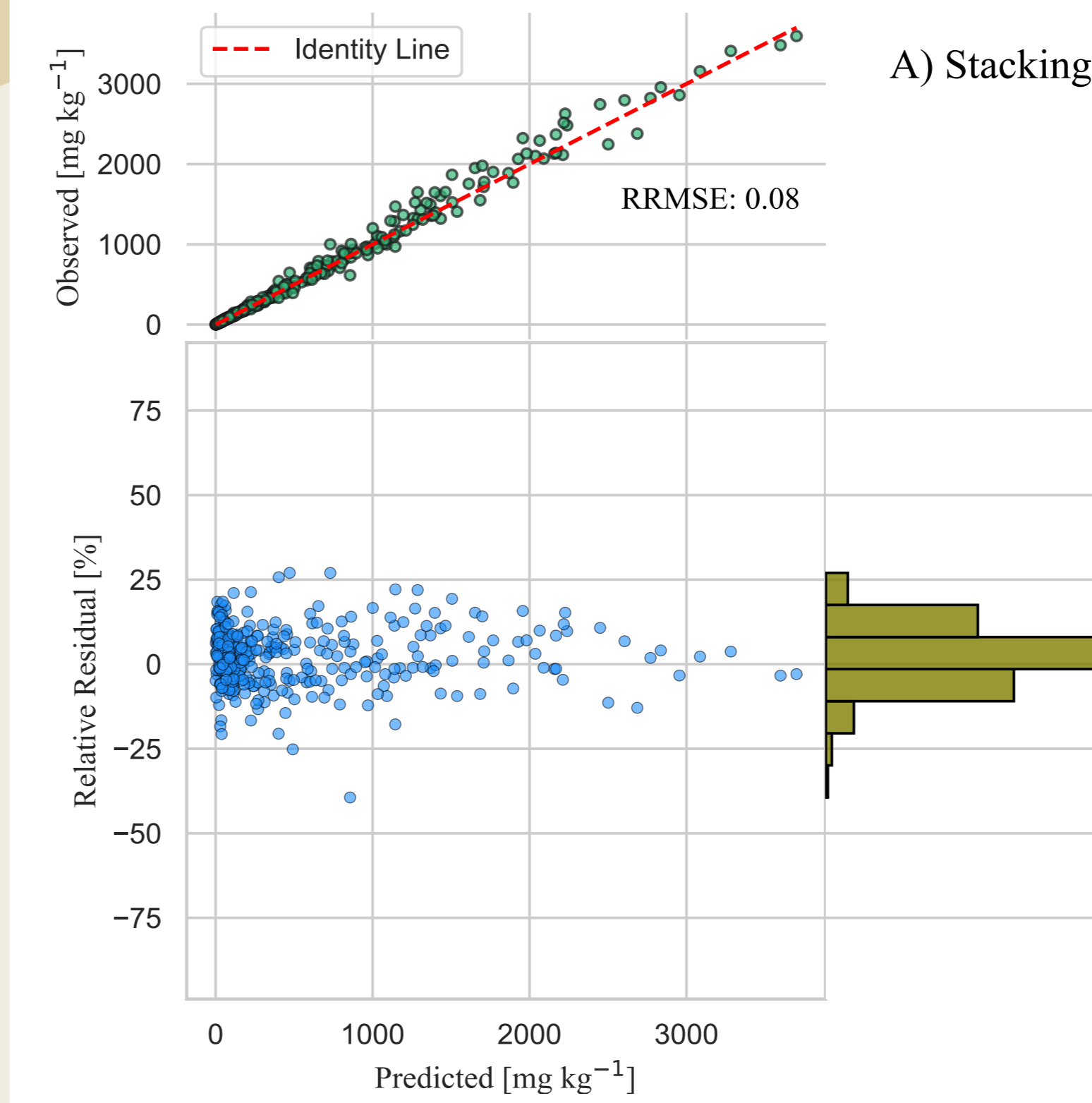
Physics-based model: Advective-Dispersive IntraParticle Pore Diffusion (ADE-IPPD) model



- Process-based numerical modeling of chemical compounds' release from soils in column leaching tests requires significant computational resources.
- Optimized and trained 200 surrogate models based on different Virtual-Reality Datasets (VRD).
- VRD is simulated data based on Halton Sequence Sampling: comprehensive parameter range coverage.
- Ensemble surrogate models: stacking and Weighted-Distance Average (WDA) methods, speed up model calibration process.

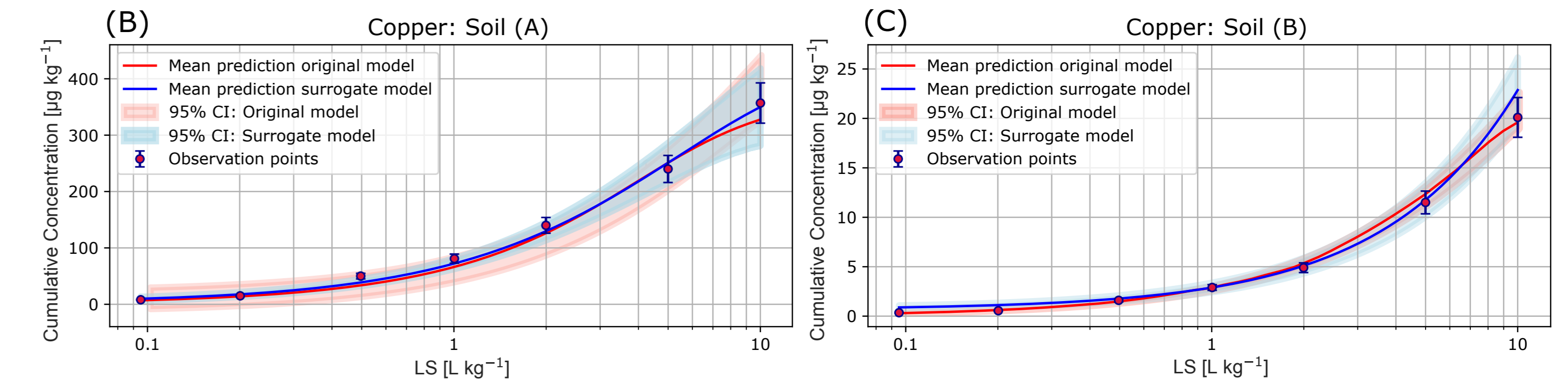
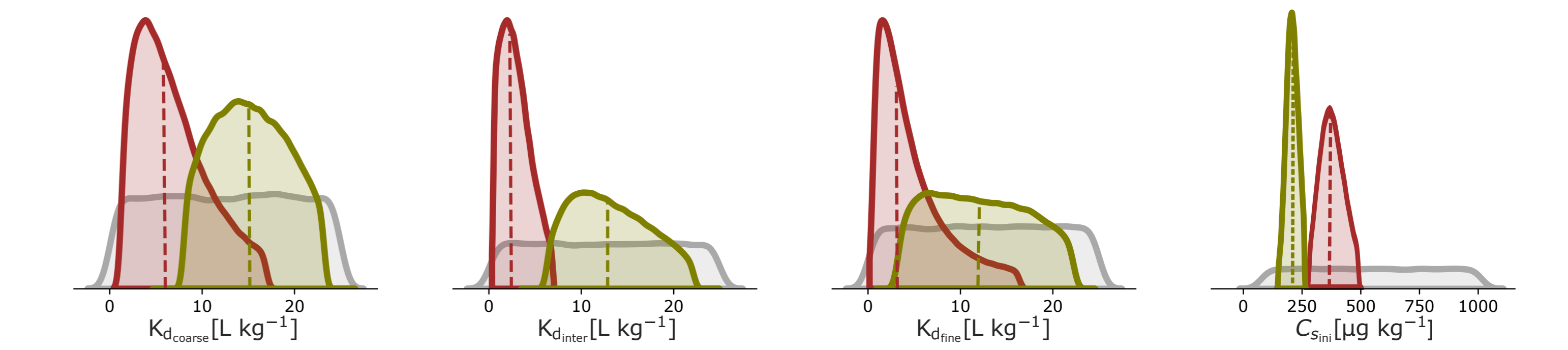
SURROGATE MODEL ACCURACY

- Stacking model outperforms WDA by non-linearly combining base-models (Extratrees) and correcting errors through meta-model (Random forest) training.
- Prioritizing maximum Mahalanobis distances in stacking from each base-model to the previous stacked-model sped up convergence through enhanced exploration and diverse region learning.

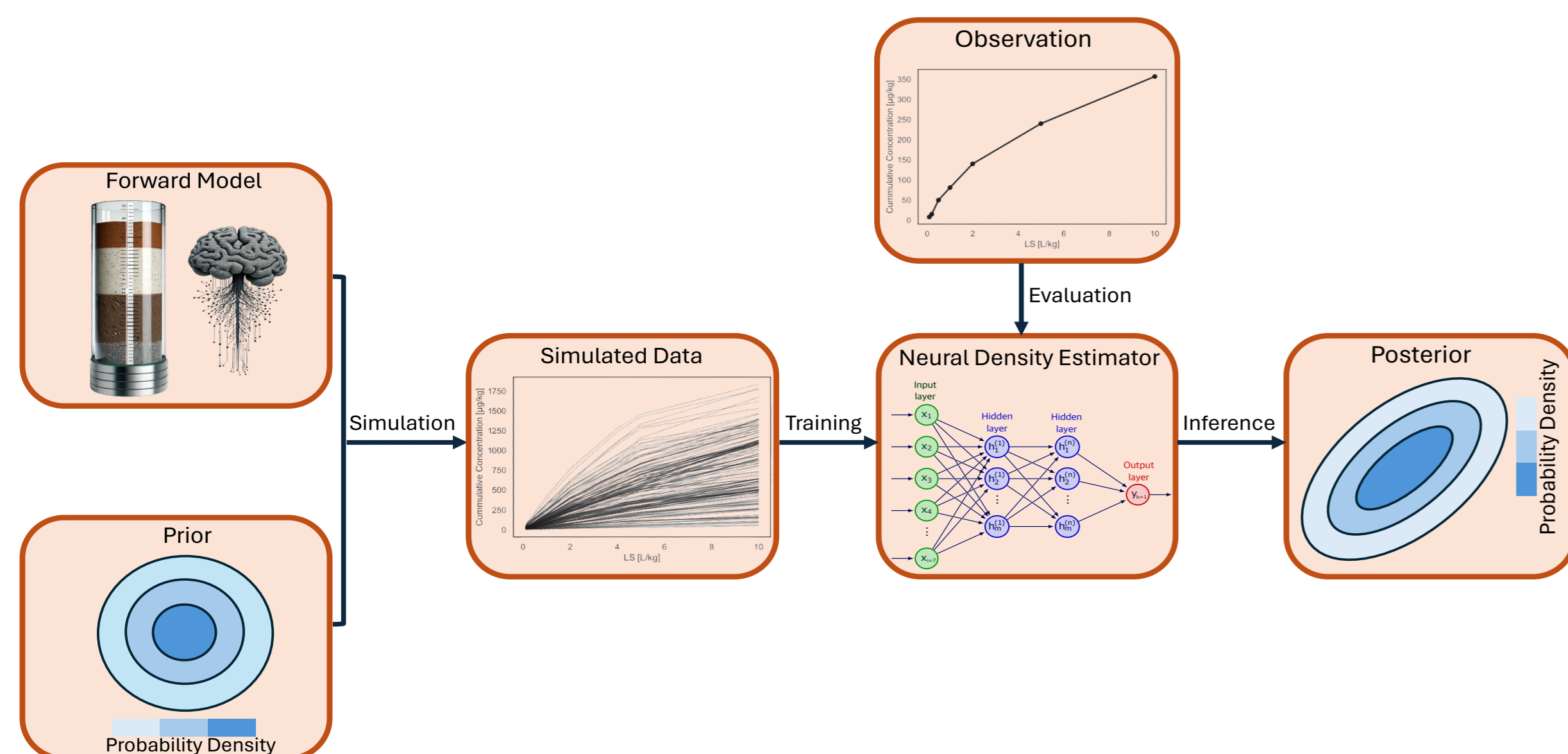


MODEL CALIBRATION

- Parameter estimation shows the highest sensitivity for C_{sini} , then K_d , based on experimental data.
- Correlations among parameters, like K_d , widen their posterior distributions, increasing uncertainty.
- Parameters from the posterior show perfect fits and similar sensitivity in both surrogate and ADE-IPPD models.



SIMULATION-BASED INFERENCE (SBI)



- SBI utilizes Bayesian methods and neural networks, notably Neural Posterior Estimation (NPE), for robust posterior parameter estimation from observed data.
- Simulated data generation via forward model is facilitated by a stacked-based ensemble surrogate model.
- Trained networks allow cost-effective inference for new measurements without additional simulations.

CONCLUSIONS

- Stacking-based ensemble surrogate model closely mimics ADE-IPPD model with 0.08 RRMSE.
- Accelerates evaluations 1000x: ADE-IPPD >150s, ensemble surrogate 0.1s.
- Simulation-based inference with an ensemble surrogate model perfectly matches experimental data using posterior distributions obtained from NPE.

