The paper's numerical experiments can be divided into six parts. Each experiment is in its specific folder.

Regularized vs Unregularized computational cost

reg vs unreg.ipynb: This notebook generates two sets of 1D samples from normal distributions with varying support sizes, then calculates the unregularized and regularized Wasserstein distance between these sets. Computational times for both regularized and unregularized Wasserstein distances are recorded in separate .txt files.

Draw Figures.ipynb: This notebook reads the computational times from the .txt files and plots a comparison of the regularized and unregularized Wasserstein distance computation times.

Note: The regularized.txt and unregularized.txt files contain the computational times from a prior experiment. Running reg vs unreg.ipynb will overwrite these files.

Agglomerative vs K-means computational cost

Data Generation.ipynb: This notebook generates samples from bivariate normal distributions and writes them into .txt files. Ten such files are generated, with the number of systems and the support size of distributions controlled via num_lists and num_tuples, respectively.

Agglomerative_comparison.ipynb: This notebook reads the generated data and applies the proposed agglomerative clustering algorithm, storing the computational times in a list.

K_means.ipynb: This notebook applies the k-means clustering algorithm to the generated distributions, recording the computational times.

Draw Figures.ipynb: This notebook uses the .txt files containing computational times to plot comparisons.

Note: The files Kmeans_systems.txt and Agglomerative_systems.txt record computational times for k-means and agglomerative clustering when varying the number of systems with fixed support sizes. Similarly, Kmeans_support.txt and Agglomerative_support.txt contain computational times for both clustering methods when varying support sizes with fixed numbers of systems.

Fixed Number Experiment

Generate Staffing Configurations.ipynb: This notebook generates the staffing configurations used in the simulation model.

Fixed Number Simulation.ipynb: This notebook simulates each staffing configuration over a fixed number of replications, writing the outputs as a list of lists to my_list.txt.

Note: Running this file will overwrite the simulation data used for the clustering algorithm.

agglomerative clustering fixed number of operators.ipynb: This notebook reads my_list.txt and performs clustering on the empirical distributions. The output includes several plots used in the paper, such as the dendrogram, silhouette index plot, correlation plots, marginal PDFs for barycenters, and marginal CDFs for staffing configurations within cluster 4.

3d simplex.ipynb: This notebook generates a 3D simplex of all simulated staffing configurations, projected onto a 2D surface, with selected configurations highlighted.

Fixed Cost Experiment

Generate Staffing Configurations.ipynb: This notebook generates the staffing configurations used in the simulation model.

Fixed Cost Simulation.ipynb: This notebook simulates each staffing configuration over a fixed number of replications, writing the outputs as a list of lists to my_list_cost.txt.

Note: Running this file will overwrite the simulation data used for the clustering algorithm.

agglomerative clustering fixed cost.ipynb: This notebook reads my_list_cost.txt and performs clustering on the empirical distributions. The output is the marginal PDFs for barycenters.

Online Monitoring

Call Center state storing.ipynb: This notebook simulates the call center over 5000 days, recording states and associated statistics. Staffing configurations are saved in state_inputs.txt, and the corresponding output distributions are stored in state_outputs.txt. An additional day of simulation logs all state changes along with their timestamps in state_inputs_for_monitoring.txt and state_times_for_monitoring.txt.

Note: Running this notebook will overwrite files used for clustering.

agglomerative clustering online monitoring.ipynb: This notebook clusters the empirical distributions of observed states and plots the marginal PDFs of the cluster barycenters. It then classifies states observed on an additional day and visualizes the state distribution for that day.

CRN

Data Generation Independent.ipynb: This notebook independently generates data across systems.

Data Generation CRN.ipynb: This notebook generates data with correlations across systems.

Data Generation True.ipynb: This notebook simulates each staffing configuration with 2000 replications to produce output distributions regarded as true distributions.

Note: Running any of these notebooks will overwrite the data used in the experiments.

agglomerative clustering CRN.ipynb: This notebook performs clustering on 100 sets of distributions, both for independent and correlated sampling, and stores the results in cluster_output_#.txt files. It also calculates the regularized Wasserstein distance between two specific staffing configurations under both sampling methods.

Calculate ARI.ipynb: This notebook uses clustering results to calculate the Adjusted Rand Index (ARI), generating a histogram of ARI values for both independent and correlated sampling scenarios.

Note: Because of the instructions of IJDS, we have put all of the .txt files in the “data” folder, but to help the user understand the code better, we have also put the .txt files in the folder of each experiment.

Note: The codes have been designed in a way that they save the figures in the “results” folder.

Note: The simulation experiments are expensive to run; thus, if the simulations are being conducted on Ocean Code, it should be considered due to the computational limitations of this website.

Note: We have uploaded the data used to produce our results in the data_backup folder to ensure it is preserved in case the files in the data folder are overwritten when running the simulation codes.

.