Design and Analysis of Computer Experiments with both Numeral and Distributional Inputs

Chunya Li, Xiaojun Cui, and Shifeng Xiong

ABSTRACT
Nowadays stochastic computer simulations with both numeral and distributional inputs are widely used to mimic complex systems which contain a great deal of uncertainty. This article studies the design and analysis issues of such computer experiments. First, we provide preliminary results concerning the Wasserstein distance in probability measure spaces. To handle the product space of the Euclidean space and the probability measure space, we prove that, through the mapping from a point in the Euclidean space to the mass probability measure at this point, the Euclidean space can be isomorphic to the subset of the probability measure space, which consists of all the mass measures, with respect to the Wasserstein distance. Therefore, the product space can be viewed as a product probability measure space. We derive formulas of the Wasserstein distance between two components of this product probability measure space. Second, we use the above results to construct Wasserstein distance-based space-filling criteria in the product space of the Euclidean space and the probability measure space. A class of optimal Latin hypercube-type designs in this product space are proposed. Third, we present a Wasserstein distance-based Gaussian process model to analyze data from computer experiments with both numeral and distributional inputs. Numerical examples and real applications to a metro simulation are presented to show the effectiveness of our methods.

ARTICLE HISTORY
Received April 2022
Accepted January 2023

KEYWORDS
Gaussian process model; Metro simulation; Mixed inputs; Space-filling design; Wasserstein distance

1. Introduction
Nowadays more and more real systems can be studied virtually by means of computer codes. Since it is often time-consuming to run such codes, elaborate design and modeling for computer experiments are necessary. A great amount of literature has studied statistical issues related to computer experiments, including experimental design, surrogate model, optimization, and many others (Fang, Li, and Sudjianto 2005; Santner, Williams, and Notz 2018). Most existing research is focused on deterministic computer experiments, that is, the simulator produces the same result if run twice using the same set of inputs. Simulations for physical phenomena through numerically solving mathematical models (differential equations) are usually deterministic. However, many real complex systems contain a great deal of uncertainty, and stochastic computer simulations can better mimic them (Baker et al. 2022). If run twice using the same set of inputs, a stochastic computer simulation yields different results. In other words, the output is a random variable. Typical stochastic computer simulations can be found in urban transportation (Elefteriadou 2014) since the movements of vehicles, pedestrians, and passengers have high degree of randomness. Other examples of stochastic computer simulations appear in physical phenomenon simulations with random inputs (Xiu 2010), reliability (Nanty et al. 2016), social science (Squazzoni, Jager, and Edmonds 2014), epidemiology (Ozik et al. 2021), and many other areas (Baker et al. 2022).

The background of this work is a metro passenger flow simulation, but our methods can be applied to other stochastic simulations. Here we introduce the simulation. Urban metro systems are important components of urban transportation systems. The implementation of a metro system simulation provides a powerful instrument for system performance monitoring, which enables operators to characterize the level of service and make decisions accordingly (Mo et al. 2021). Such a simulation is a typical stochastic simulation since a metro system contains a great deal of passengers’ uncertainty.

The simulation we consider in this article is designed to mimic passenger flows on a typical metro route. On this route a passenger begins to tap in at the origin station and ends in tapping out at the destination station with taking only a train. The time between the arrival and departure of a passenger can be divided into access time, wait time, time on board, and egress time. Access time is the time it takes the passenger to walk from the tap-in fare gate to the platform; wait time is the time for which the passenger waits on the platform until boarding a train; and egress time is the time it takes to walk to the tap-out fare gate after alighting from the train.

During peak hours, passengers may miss one or more trains due to crowded platform and carriages, and there may be more...
than one possible itineraries between their tap-in and tap-out times. Figure 1 illustrates all possible itineraries for passenger $i$ who enters and exits the metro system; a similar figure can be found in Zhu, Koutsopoulos, and Wilson (2017).

We can use the passenger flow simulation to study the influence of the number of passengers, train schedule, and other factors on crowdedness, passenger-to-train assignment, and other responses we are interested in. Specifically, for a one-day simulation, the inputs are as follows.

(i) number of passengers in the day;
(ii) each passenger’s origin and destination stations and tap-in time;
(iii) boarding probability of a passenger at each station (except the last station);
(iv) probability distribution of a passenger’s access time at each station (except the last station);
(v) probability distribution of a passenger’s egress time at each station (except the first station);
(vi) train capacity;
(vii) train schedule including arrival and departure times at each station.

The simulation can yield movement of each passenger in the metro system, including which train he/she takes, his/her tap-out time, and others. Clearly, it is a stochastic simulation, including both numeral and distributional inputs. In real cases, many stations have more than one accesses and exits, which may result in complex distributions of walking time such as a mixture of several uni-modal distributions (Xiong et al. 2022). Figure 2 shows two density functions of egress time at a station in Beijing, China. It is hard to represent the distributional inputs in (iv) and (v) with known parametric distributions. In metro emergencies, these distributions are more difficult to predict and to describe. Therefore, we cannot view these inputs as Euclidean parameters, and need to use a infinite-dimensional probability measure space as their domain.

Like the above metro simulation, the randomness of the output of a stochastic simulation is caused by some stochastic mechanism involved in the computer codes. Probability
distributions corresponding to the stochastic mechanism can usually be viewed as distributional inputs. There are a few articles on the design and modeling issues of computer experiments with function inputs (Muehlenstaedt, Fruth, and Roustant 2017; Tan 2019; Betancourt et al. 2020; Chen et al. 2021), but very limited on distributional inputs. Bachoc et al. (2018) constructed the Wasserstein distance-based Gaussian process model with a one-dimensional distributional input. Bachoc et al. (2020) discussed such a model with multidimensional distributional inputs. The two articles do not consider the experimental design issue of distributional inputs. Note that actual stochastic simulations often have mixed types of inputs like the above metro simulation. In this article we focus on computer experiments with both numeral and distributional inputs. To the best of our knowledge, the design and analysis issues of such computer experiments have not been investigated in the literature.

In a probability measure space, the Wasserstein distance, which is related to the optimal transport problem (Panaretos and Zemel 2020), possesses a number of good mathematical properties. It thus has been widely applied in machine learning and statistics (Arjovsky, Chintala, and Bottou 2017; Peyré and Cuturi 2019), including the construction of surrogate models of computer simulations mentioned above. For the mixed-input problem we face, we find that it is also very suitable. We prove that, through the mapping from a point in the Euclidean space to the mass probability measure at this point, the Euclidean space can be isomorphic to the subset of the probability measure space, which consists of all the mass measures, with respect to the Wasserstein distance. Therefore, the product space of the Euclidean space and the probability measure space can be viewed as a product probability measure space. By this way, the mixed inputs are unified within this product probability measure space. We derive formulas of the Wasserstein distance between two components of the product space, and use it to define Wasserstein distance-based space-filling criteria. A class of discretization-based approximate maximin designs in a one-dimensional probability measure space and a class of Latin hypercube-type designs in the product space are proposed. For the modeling issue, we focus on real-valued responses which are numeral features such as the expectation of the output distribution. We use the Wasserstein distance to construct a Gaussian process model with mixed inputs, and discuss the corresponding estimation and prediction methods. Numerical experiments with test functions are presented to evaluate our design and prediction methods. Real applications to the metro passenger flow simulation mentioned previously are provided. We apply the proposed methods to build the surrogate model of the simulation, which shows how passengers’ travel times depend on their walking time distribution and a boarding probability parameter.

The rest of this article is organized as following. Section 2 gives preliminary results concerning the Wasserstein distance. Section 3 constructs experimental designs in the product space of the Euclidean space and the probability measure space. Section 4 discusses the Gaussian process model for the mixed inputs. Section 5 shows numerical results with test functions, Section 6 provides real applications to the metro simulation. Section 7 concludes the article with a discussion. Technical proofs and additional numerical results are given in the supplementary materials.

2. The Wasserstein Distance

For a vector \( \mathbf{x} = (x_1, \ldots, x_d)' \in \mathbb{R}^d \), let \( \| \mathbf{x} \|_p = (|x_1|^p + \cdots + |x_d|^p)^{1/p} \) denote its \( \ell_p \) norm, where \( ' \) represents transpose and \( p \geq 1 \). For \( p, q \geq 1 \), consider the set \( \mathcal{P}_{q,p}(\mathbb{R}^d) \) of probability measures on \( \mathbb{R}^d \) with a finite moment of order \( q \) with respect to the \( \ell_p \) norm, that is, \( \mu \in \mathcal{P}_{q,p}(\mathbb{R}^d) \) if

\[
\int \| \mathbf{x} \|_p^q d\mu(\mathbf{x}) = \int (|x_1|^p + \cdots + |x_d|^p)^{q/p} d\mu(x_1, \ldots, x_d) < \infty. \tag{1}
\]

For \( \mu, \nu \in \mathcal{P}_{q,p}(\mathbb{R}^d) \), we denote by \( \Pi(\mu, \nu) \) the set of all probability measures \( \pi \) over the product set \( \mathbb{R}^d \times \mathbb{R}^d \) with first (resp. second) marginal \( \mu \) (resp. \( \nu \)). Any element in \( \Pi(\mu, \nu) \) is called a coupling measure of \( \mu \) and \( \nu \). With the cost function \( c(\mathbf{x}, \mathbf{y}) = \| \mathbf{x} - \mathbf{y} \|_p^q \), the transportation cost between \( \mu \) and \( \nu \) is defined as

\[
\mathcal{T}_{q,p}(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int c(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}) = \inf_{\pi \in \Pi(\mu, \nu)} \int \| \mathbf{x} - \mathbf{y} \|_p^q d\pi(\mathbf{x}, \mathbf{y}).
\]

Note that \( \mathbb{R}^d \), equipped with \( \ell_p \) norm, is a Polish space, and that \( \| \mathbf{x} - \mathbf{y} \|_p^q \geq 0 \). The above infimum can be reached (Theorem 4.1 of Villani 2009). The Wasserstein distance (also called Monge-Kantorovich distance) with ground metric \( \ell_p \) between \( \mu \) and \( \nu \) is defined as

\[
W_{q,p}(\mu, \nu) = (\mathcal{T}_{q,p}(\mu, \nu))^{1/q}.
\]

When \( d = 1 \), \( \mathcal{P}_{q,p}(\mathbb{R}) \), \( \mathcal{T}_{q,p}(\mu, \nu) \), and \( W_{q,p}(\mu, \nu) \) are independent of \( p \) by (1), and we denote them by \( \mathcal{P}_q(\mathbb{R}) \), \( T_q(\mu, \nu) \), and \( W_q(\mu, \nu) \), respectively. For this case, the Wasserstein distance can be computed by (Villani 2009)

\[
W_q(\mu, \nu) = \left\{ \int_0^1 \left[ F_{\mu}^{-1}(t) - F_{\nu}^{-1}(t) \right]^{q-1} dt \right\}^{1/q}, \tag{2}
\]

where \( F_{\mu} \) and \( F_{\nu} \) are the cumulative distribution functions of \( \mu \) and \( \nu \), respectively, and \( F_{\mu}^{-1}(t) = \inf \{ u : F_{\mu}(u) \geq t \} \), \( F_{\nu}^{-1}(t) = \inf \{ u : F_{\nu}(u) \geq t \} \).

Let \( \delta_x \) denote the mass probability measure at \( \mathbf{x} \in \mathbb{R}^d \), and \( \Delta^d = \{ \delta_x : \mathbf{x} \in \mathbb{R}^d \} \). Consider the product space \( \Delta^d \times \mathcal{P}_q(\mathbb{R}) \), which is a subset of \( \mathcal{P}_{q,p}(\mathbb{R}^{d+1}) \) with any \( p \geq 1 \).

**Lemma 1.** For \( \mathbf{x}, \mathbf{y} \in \mathbb{R}^d \), we have

\[
W_{q,p}(\delta_x, \delta_y) = \| \mathbf{x} - \mathbf{y} \|_p.
\]

This lemma indicates that, with the mapping \( \mathbf{x} \mapsto \delta_x \), \( \mathbb{R}^d \) (equipped with the \( \ell_p \) distance) is isomorphic to \( \Delta^d \) with respect to the Wasserstein distance. Therefore, the product space \( \mathbb{R}^d \times \mathcal{P}_q(\mathbb{R}) \) is isomorphic to \( \Delta^d \times \mathcal{P}_q(\mathbb{R}) \), which is a subspace of \( \mathcal{P}_{q,p}(\mathbb{R}^{d+1}) \). We view any combination of numeral and distributional inputs as a component of this probability measure space. This provides a way to unify such mixed inputs in \( \mathbb{R}^{d+1} \times \mathcal{P}_q(\mathbb{R}) \).

For the special product probability measure space \( \Delta^d \times \mathcal{P}_q(\mathbb{R}) \), the Wasserstein distance between its two components has the following expression.
Theorem 1. For \( x, y \in \mathbb{R}^d \), \( \mu, \nu \in \mathcal{P}_d(\mathbb{R}) \), we have
\[
W_{q,p}(\delta_x \times \mu, \delta_y \times \nu) = \left\{ \int_0^1 c(F_\mu^{-1}(t) - F_\nu^{-1}(t)) \, dt \right\}^{1/q},
\]
where \( c(z) = (\|x - y\|_p^p + |z|^p)^{\frac{q}{p}} \).

This theorem presents a formula for calculating the Wasserstein distance in the space we focus on. In particular, we provide its two important special cases,
\[
W_{1,1}(\delta_x \times \mu, \delta_y \times \nu) = \|x - y\|_1 + W_1(\mu, \nu), \quad (3)
\]
\[
W_{2,2}(\delta_x \times \mu, \delta_y \times \nu)^2 = \|x - y\|_2^2 + W_2(\mu, \nu)^2. \quad (4)
\]
In fact, (3) and (4) can be generalized to the case where \( \mu \) and \( \nu \) are defined on higher dimensional spaces. More precisely, we have the following theorem.

Theorem 2. For \( x, y \in \mathbb{R}^{d_1} \), \( \mu, \nu \in \mathcal{P}_{p,p}(\mathbb{R}^{d_2}) \) with \( d_1, d_2 \in \mathbb{N} \) and \( p \geq 1 \), we have
\[
W_{p,p}(\delta_x \times \mu, \delta_y \times \nu)^p = \|x - y\|_p^p + W_{p,p}(\mu, \nu)^p.
\]

3. Experimental Design

From the results in the previous section, the Wasserstein distance can be viewed as an extension of the \( \ell_p \) distance in the Euclidean space to the probability measure space. Similar to the Euclidean space, we can define the Wasserstein distance-based space-filling designs in the probability measure space. In this section we first present a method to construct space-filling designs in the space of probability measures in one dimension. We then introduce Latin Hypercube (LH)-type space-filling designs for the product space of the Euclidean space and the probability measure space.

3.1. Space-Filling Designs in \( \mathcal{P}([0, 1], \tau) \)

Note that distributional inputs of stochastic computer simulations usually have compact support and certain degrees of smoothness. We consider the space \( \mathcal{P}([0, 1], \tau) = \{ \mu \in \mathcal{P}_d(\mathbb{R}) : \text{support of } \mu \subset [0, 1], \sup_{x,y \in [0,1]} |F_\mu(x) - F_\mu(y)|/|x - y| \leq \tau \} \) for some constant \( \tau > 0 \). If \( F_\mu \) has a density \( f_\mu \), then \( \tau \) can be set as \( \text{max}_{x \in [0,1]} f_\mu(x) \).

For a design of \( n \) runs, \( D = \{ \mu_1, \ldots, \mu_n \} \subset \mathcal{P}([0, 1], \tau) \), the minimum Wasserstein distance criterion is
\[
\text{mdc}(D) = \min_{1 \leq i < j \leq n} W_q(\mu_i, \mu_j).
\]

Note that the parameter \( q \) in the Wasserstein distance can be arbitrary due to finite support. We can compute \( W_q(\mu_i, \mu_j) \) by (2). The design that maximizes the criterion (5) is defined as the maximin Wasserstein distance design. Similarly we can define minimax Wasserstein distance design and minimum \( \phi_0 \) design, which are analogues of those in Euclidean space (Johnson, Moore, and Ylvisaker 1990; Morris and Mitchell 1995).

Since \( \mathcal{P}([0, 1], \tau) \) is an infinite-dimensional space, the problem in optimizing (5) is difficult. We use a discretization method to approximate it. For \( \mu \in \mathcal{P}([0, 1], \tau) \), \( F_\mu \) can be approximated by the piecewise linear function
\[
\hat{F}_\mu(x) = \sum_{i=1}^{m-1} \left\{ F_\mu \left( \frac{i-1}{m-1} \right) + (m-1) \left( x - \frac{i-1}{m-1} \right) \right\} I \left( x \in \left[ \frac{i-1}{m-1}, \frac{i}{m-1} \right), \right.
\]
where \( I \) is the indicator function. Therefore, \( \mathcal{P}([0, 1], \tau) \) can be approximated by the finite dimensional space
\[
\mathcal{P}_{m-1}([0, 1], \tau) = \left\{ \mu : F_\mu(x) = \sum_{i=1}^{m-1} \frac{s_i + (m-1)(x - \frac{i-1}{m-1})}{m-1} (s_{i+1} - s_i) \right\} \quad \text{for all } \left( s_1, \ldots, s_m \right)\text{ with } s_0 = 0 = s_{m+1} \leq \cdots \leq s_m = 1, s_{i+1} - s_i \leq \tau/(m-1), i = 1, \ldots, m-1.
\]

Each \( \mu \in \mathcal{P}_{m-1}([0, 1], \tau) \) corresponds to the vector \( (s_1, \ldots, s_m)' \) of \( m - 2 \) degrees of freedom. Let \( t_i = s_{i+1} - s_i, i = 1, \ldots, m-1 \). Then \( \mu \) is specified by the vector \( (t_1, \ldots, t_{m-1})' \in E_{m-1}(\tau) = \{ (x_1, \ldots, x_{m-1})' : 0 \leq x_i \leq \tau/(m-1), i = 1, \ldots, m-1, \sum_{i=1}^{m-1} x_i = 1 \} \). The maximin Wasserstein distance design problem reduces to
\[
\max \xi(t_1, \ldots, t_n) = \min_{1 \leq i < j \leq n} W_q(\mu_i, \mu_j) \quad \text{subject to } t_1, \ldots, t_n \in E_{m-1}(\tau),
\]
where \( \mu_i \) corresponds to \( t_i \) for \( i = 1, \ldots, n \). Let \( \{ t_1', \ldots, t_n' \} \) denote the solution to (8). Then the corresponding \( \{ \mu_1', \ldots, \mu_n' \} \) can be viewed as an (approximate) maximin Wasserstein distance design in \( \mathcal{P}([0, 1], \tau) \).

There are \( n \times (m - 1) \) variables we need to optimize in the approximate problem (8), which is still a high-dimensional
optimization problem. We use the block coordinate descent algorithm (Tseng 2001) to solve (8). In each iteration we only optimize one run in the design. See Algorithm 1 for detailed steps. This algorithm can also be used to construct approximate optimal designs with other criteria, such as the minimax Wasserstein distance design and distance-based space-filling. Therefore, \( D \) can be computed by Theorem 1.

Figure 3 shows the density functions in six-run maximin distance-based designs in this space. For instance, similar to (8), the maximin Wasserstein distance design problem is
\[
\max_{\mu} \mathcal{E}(\mu) = \min_{\mu, \delta, \mu_j} W_{q,p}(\delta_{\mu} \times \mu_j, \delta_{\mu} \times \mu_j)
\]
subject to \( \mu \in P([0,1], \tau) \). The projections of \( D(\varepsilon_1, \ldots, \varepsilon_n) \) on \([0,1]^d \) and \( P([0,1], \tau) \) are respectively \( D_1 \) and \( D_2 \), which are both space-filling. Therefore, \( D(\varepsilon_1, \ldots, \varepsilon_n) \) can be viewed as an LH-type design (Mckay, Beckman, and Conover 1979). The designs \( D_1 \) and \( D_2 \) can be called base designs of \( D(\varepsilon_1, \ldots, \varepsilon_n) \). We can set \( D_2 \) as the maximin Wasserstein distance design proposed in the previous section. For \( D_1 \), there are many feasible choices (Joseph 2016; Santner, Williams, and Notz 2018). Here we prefer to those with good projection properties such as the maximin LH design (Park 1994), the maximum projection design (Joseph, Gul, and Ba 2015; Mu and Xiong 2018), and the rotated sphere packing design (He 2017), which usually yield good prediction of the simulation output.

Similar to optimal LH designs, we define the maximin Wasserstein distance LH-type design based on \((D_1, D_2)\) is the set \( D^* = D(e_1^*, \ldots, e_n^*) = \{(\mu_1, \mu_1^*), \ldots, (\mu_n, \mu_n^*)\} \) of \( [0,1]^d \) and \( P([0,1], \tau) \), where \( e_1^*, \ldots, e_n^* \) maximizes
\[
\min_{1 \leq i < j \leq n} W_{q,p}(\delta_{\mu_1} \times \mu_1, \delta_{\mu_j} \times \mu_j)
\]
over all permutations \( (e_1, \ldots, e_n) \) of \( (1, \ldots, n) \). The Wasserstein distance in (10) can be computed by Theorem 1, especially by (3) and (4). Usually it is infeasible to compute the global solution.
to maximize (10) based on all the $n!$ permutations. We can use the Monte Carlo method to approximate it by generating a large number of random permutations. Figure 5 shows the maximin $W_{2,2}$ distance LH-type design and the maximin $W_{1,1}$ distance LH-type design in $[0, 1] \times \mathcal{P}([0, 1], 3)$ for $n = 6$, where the base designs are the uniformly scattered points $\{0, 1/5, 2/5, \ldots, 1\}$ and the corresponding maximin Wasserstein distance designs in Figure 3. Here the measures plotted with the red solid line, red
dotted line, blue solid line, blue dotted line, green solid line, and green dotted line in Figure 3 are denoted by $\mu_1, \ldots, \mu_6$. From Figure 5 we can see that the optimal permutations in the two LH-type designs are $(5,3,6,1,4,2)$ and $(5,3,2,6,1,4)$, respectively.

We can use the above maximin Wasserstein distance LH-type design as the initial point of the algorithm for constructing maximin Wasserstein distance design defined in (9). The two types of designs will be compared in Section 5.

### 4. Gaussian Process Modeling

This section builds a Gaussian process model for computer experiments with both numerical and distributional inputs. We first define $h \circ \mu$ to be the probability measure of the random variable $h(X)$ for a function $h: \mathbb{R} \rightarrow \mathbb{R}$, where $X \sim \mu$. For $x \in [0,1]^d$ and $\mu \in \mathcal{P}_2(\mathbb{R})$, we model the output of such a computer simulation as

$$f(x, \mu) = g(x)' \beta + \int \alpha(t)' d(h \circ \mu(t)) + Z(x, \mu),$$

where $g(\cdot) = (g_1(\cdot), \ldots, g_n(\cdot))'$ and $h(\cdot) = (h_1(\cdot), \ldots, h_n(\cdot))'$ are pre-specified functions, $h \circ \mu = (h_1 \circ \mu, \ldots, h_n \circ \mu)$, $\beta$ is a vector of unknown regression coefficients, $\alpha(\cdot)$ is a vector of unknown smooth functions, and $Z(x, \mu)$ is a stationary Gaussian process defined on $[0,1]^d \times \mathcal{P}_2(\mathbb{R})$ with mean zero, variance $\sigma^2$, and covariance structure given below. The covariance between $Z(x, \mu_1)$ and $Z(x, \mu_2)$ in (11) is represented by

$$\text{cov}[Z(x_1, \mu_1), Z(x_2, \mu_2)] = \sigma^2 R(x_1, x_2; \mu_1, \mu_2 | \theta),$$

where $R$ is the Gaussian correlation function

$$R(x_1, x_2; \mu_1, \mu_2 | \theta) = \exp \left\{ - \sum_{i=1}^d \theta_{i}(x_{1i} - x_{2i})^2 - \theta_{d+1} W_{2,2}(\mu_1, \mu_2)^2 \right\},$$

(12)

with positive correlation parameters $\theta \in \mathbb{R}^d$. Note that $R_W(\mu_1, \mu_2) = \exp \left\{ -\theta_{d+1} W_{2,2}(\mu_1, \mu_2)^2 \right\}$ gives a valid correlation structure of Gaussian processes on $\mathcal{P}_2(\mathbb{R})$ (Bachoc et al. 2018, 2020). The correlation structure in (12) is valid for Gaussian processes on the product space of $[0,1]^d$ and $\mathcal{P}_2(\mathbb{R})$. It should be noted that the Gaussian process models proposed by Bachoc et al. (2018, 2020) are constructed for only distributional inputs, and that they do not use linear regression terms to describe the overall trend of the response.

Since the infinite-dimensional $\alpha(\cdot)$ in (11) is hard to estimate, we set $\zeta = 1$ in the following, and thus only one function $\alpha(\cdot)$ is unknown. We use a linear expression to parameterize it. Let

$$\alpha(\cdot) = y' b(\cdot)$$

(13)

with $y = (y_1, \ldots, y_n)' \in \mathbb{R}_n$, where $b(\cdot) = (b_1(\cdot), \ldots, b_l(\cdot))'$ are pre-specified basis functions. Then (11) can be approximated as

$$f(x, \mu) = g(x)' \beta + \left\{ \int b(t)d(h \circ \mu(t)) \right\}' y + Z(x, \mu).$$

(14)

The parameters in model (14) can be estimated by the maximum likelihood method. Suppose the set of input values is $\{(x_1, \mu_1), \ldots, (x_n, \mu_n)\} \subset [0,1]^d \times \mathcal{P}_2(\mathbb{R})$. The corresponding response values are $y = (f(x_1, \mu_1), \ldots, f(x_n, \mu_n))'$. The negative log-likelihood, up to an additive constant, is proportional to

$$n \log(\sigma^2) + \log(\det(R)) + (y - G \beta - J'y)' R^{-1} (y - G \beta - J'y)/\sigma^2,$$

(15)

where $R$ is the $n \times n$ correlation matrix whose $(i, j)$th entry is $R(x_i, x_j; \mu_i, \mu_j | \theta)$ defined in (12), “det” denotes matrix determinant, $G = (g(x_1), \ldots, g(x_n))'$, and $J = (\int b(t)d(h \circ \mu_1(t)), \ldots, \int b(t)d(h \circ \mu_n(t)))'$. Denote $U = (G)'$ and $\psi = (\beta' y')'$.

When $\theta$ is known, the maximum likelihood estimators (MLEs) of $\psi$ and $\sigma^2$ are

$$\hat{\psi} = (U'R^{-1}U)^{-1}U'R^{-1}y,$$

$$\hat{\sigma}^2 = (y - U\hat{\psi})'R^{-1}(y - U\hat{\psi})/n.$$
where $r_0 = \{(X_0, x_j; \mu_0, \mu_1 \mid \theta), \ldots, (X_0, x_j; \mu_0, \mu_n \mid \theta)\}'. Clearly this predictor possesses the interpolation property. Given $\kappa \in (0, 1)$, the $100(1-\kappa)$% prediction interval of $f(x_0, \mu_0)$ can be given by

$$P \left( f(x_0, \mu_0) \in \hat{f}(x_0, \mu_0) \pm \eta(x_0, \mu_0) t_{n-s-l}(\kappa/2) \right) = 1 - \alpha,$$

where $\eta(x_0, \mu_0) \geq 0$, and $t_{n-s-l}(\kappa/2)$ is the upper $\kappa/2$ quantile of the Student's $t$-distribution with $n - s - l$ degrees of freedom. When $\theta$ is unknown, by plugging (16) into (15), we have the MLE of $\theta$

$$\hat{\theta} = \arg \min_{\theta} n \log(\hat{\sigma}^2) + \log(\det(R)).$$

The predictor $\hat{f}$ in (17) and the prediction interval in (18) can be modified by replacing $\theta$ with $\hat{\theta}$.

Similarly to the Gaussian process model on the Euclidean space in the literature, we call the above methods without and with the linear regression terms $g(\cdot)$ and $h(\cdot)$ in (11) simple Kriging and universal Kriging, respectively. Usually we set $h$ in (11) as the identity function.

There are some methods to parameterize $\alpha(\cdot)$ in (13) such as the spline approximation (De Boor 1978). Here we propose the reconstruction parameterization approach (Xiong 2021) because of its good interpretation of the parameters. Specifically, in this approach the parameters $\gamma_i = \alpha(a_i), i = 1, \ldots, l$, where $\{a_1, \ldots, a_l\}$ is the set of knots, and $b(\cdot) = (b_1(\cdot), \ldots, b(l)(\cdot))$ in (13) are specific interpolation basis functions. When the distributional input in model (11) has a support $[0, 1]$, we select $b$ as the polynomial interpolation basis functions (De Boor 1978), which have the Lagrange forms

$$b_j(t) = \prod_{1 \leq k \leq l, k \neq j} \frac{t - a_k}{a_j - a_k}, \quad j = 1, \ldots, l.$$  

We use the Chebyshev nodes

$$A = \{a_j = 1/2 - \cos((2j - 1)\pi/2l)/2 : j = 1, \ldots, l\}$$

To avoid Runge’s Phenomenon (De Boor 1978). The number of knots can be selected as $l = 10$ according to the common 10d rule (Loepky, Sacks, and Welch 2009).

When the distributional input $\mu$ in (11) lies in $\mathcal{P}([0, 1], \tau)$, we use the design $\{\mu_1, \ldots, \mu_n\} \subset \mathcal{P}_{m-1}([0, 1], \tau)$ constructed in Section 3. Let $h$ in (11) be the identity function. For $i = 1, \ldots, n$, $j = 1, \ldots, l$, by (6) and (7), the entries of $J$ in (15) can be given by

$$\int b_j(t) dF_{\mu_i}(t) = \sum_{i=1}^{m-1} \int_{(i-1)/m-1}^{i/m-1} b_j(t) dF_{\mu_i}(t)$$

$$= (m - 1) \sum_{i=1}^{m-1} \left[ F_{\mu_i} \left( \frac{i}{m-1} \right) - F_{\mu_i} \left( \frac{i-1}{m-1} \right) \right] \int_{(i-1)/m-1}^{i/m-1} b_j(t) dt.$$  

Note that when $\mu = \delta_0$ in (11) for some $x \in \mathbb{R}$, $\int \alpha(t)'d(h \circ \mu(t))$ reduces to $\alpha(x)'h(x)$, which is consistent with the linear regression term in the numerical-input universal Kriging model (Santner, Williams, and Notz 2018). Inspired by this, (11) can be modified to model a simulation with two distributional inputs, that is, $f(\mu, \nu)$ with $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$, as

$$f(\mu, \nu) = \int \alpha(t)'d(h \circ \mu(t)) + \int \beta(t)'d(g \circ \mu(t)) + Z(\mu, \nu),$$

where $h(\cdot) = (h_1(\cdot), \ldots, h_2(\cdot))$ and $g(\cdot) = (g_1(\cdot), \ldots, g_2(\cdot))$ are pre-specified functions, $\alpha(\cdot)$ and $\beta(\cdot)$ are vectors of unknown smooth functions, and $Z(\mu, \nu)$ is a stationary Gaussian process defined on $\mathcal{P}_2(\mathbb{R}) \times \mathcal{P}_2(\mathbb{R})$ with mean zero, variance $\sigma^2$, and correlation function

$$R(\mu_1, \mu_2; \nu_1, \nu_2 | \theta, \theta_0) = \exp \left[ -\theta_0 \mu_2(\mu_1, \mu_2)^2 - \theta_0 \nu_2(\nu_1, \nu_2)^2 \right]$$

for $\theta, \theta_0 > 0$. Consequently, we can construct the Gaussian process model for computer simulation $f(x, \mu \times \nu)$ defined on $[0, 1]^d \times (\mathcal{P}_2(\mathbb{R}) \times \mathcal{P}_2(\mathbb{R})) \subset [0, 1]^d \times \mathcal{P}_{2,2}(\mathbb{R})$.

5. Numerical Experiments with Test Functions

In this section we conduct numerical experiments with the following test functions on $[0, 1]^d \times \mathcal{P}([0, 1], \tau)$,

(I) $f(x, \mu) = c + x^{1+c} + \int t d\mu(t),$  

(II) $f(x, \mu) = \int \cos(3t + c) d\mu(t) + \exp(x) + c_2 F_{\mu}(x),$  

(III) $f(x_1, x_2, \mu) = \left[ x_1 + \int t d\mu(t) + c_1 \right]^2 - c_2 \log(1 + x_2)$,  

(IV) $f(x_1, x_2, x_3, x_4, \mu) = \left[ 20x_1 + c_1 \log(10 + x_2 x_3 x_4) \right]^{-1} \int \exp[t + c_2 \sin(t)] d\mu(t),$  

(V) $f(x_1, \ldots, x_6, \mu) = x_1 + c_1 \log(10 + x_2 \cdots x_6) + c_2 F_{\mu}(1/2)^2.$

The constants $c$, $c_1$, and $c_2$ in them are generated from the uniform distribution on $[0, 1]$. Four designs and two modeling methods are compared. The four designs are as follows.

LH-L2 represents the maximin Wasserstein distance LH-type design based on $(D_1, D_2^{(2)})$ defined by (10), where $D_1$ is the maximin $\ell_2$-distance LH design and $D_2^{(2)}$ is the maximin $W_2$-distance design.
Figure 6. Root mean squared prediction errors for the test functions in Section 5.
LH-L1 represents the maximin $W_1$-distance LH-type design based on $(D_1, D_2^{(1)})$, where $D_2^{(1)}$ is the maximin $W_1$-distance design.

Mm-L2 represents the maximin $W_{2,2}$ distance design in $[0,1]^d \times \mathcal{P}([0,1], \tau)$.

Mm-L1 represents the maximin $W_{1,1}$ distance design in $[0,1]^d \times \mathcal{P}([0,1], \tau)$.

Fix $\tau = 3$. Sample sizes in our designs are set as 20, 40, 60, 80, and 100 for $d = 1, 2$ and 40, 60, 80, 100, and 120 for $d = 4, 6$. Two modeling methods include simple Kriging (SK) and universal Kriging (UK). We use $g(x) = g(x_1, \ldots, x_d) = (1, x_1, \ldots, x_d)'$ and $h(t) = t$ in the UK model (11).

For each design and each test function, we generate training data, and then use SK and UK to build prediction models $\hat{f}$ in (17). The prediction accuracy is evaluated by the empirical squared prediction error, \[ \sum_{k=1}^{N} \left\{ \hat{f}(x_\ast_k, \mu_\ast_k) - f(x_\ast_k, \mu_\ast_k) \right\}^2 / N \] with $N = 1000$, where the test data $(x_\ast_1, \mu_\ast_1), \ldots, (x_\ast_N, \mu_\ast_N) \in [0,1]^d \times \mathcal{P}([0,1], \tau)$ are generated randomly. Root mean squared prediction errors over 100 repetitions are shown in Figure 6. We can see that, with a relatively large $n$, UK is often better than SK, which is consistent with our empirical experience on Kriging in the Euclidean space. In practice, we can choose SK or UK through leave-one-out cross-validation. For the four designs, the two LH methods have better overall performance than the Mm methods, because LH possesses good projection properties. Note that we need much more computational cost in constructing Mm than LH. The proposed LH-type designs are more suitable for practice. It seems that L1 or L2 does not have clear superiority toward the other. On the computational aspect, it takes more time to compute $W_{1,1}$ than $W_{2,2}$ because $W_{2,2}$ is smoother. We recommend using LH-L2 for real applications. In addition, LH-L2 outperforms other criterion-based optimal designs; see the numerical experiment in Section C of the online supplementary materials.

6. Applications to the Metro Simulation

Let us return to the metro simulation introduced in Section 1. Our simulation aims at simulating a real metro route of six stations. The inputs (i), (ii), (v), (vi), and (vii) can be obtained or estimated from the automatic fare collection system and automatic vehicle location system (Xiong et al. 2022; Li et al. 2022). The simulator designs the boarding probability for passenger $i$ in (iii) as

$$Pr(\rho; x) = \begin{cases} 0, & \rho < 1 - x/2; \\ (\rho - 1 + x/2)/x, & \rho \in [1 - x/2, 1 + x/2]; \\ 1, & \rho > 1 + x/2, \end{cases}$$

(19)

where $\rho = (L - N_0)/N$, $L$ denotes the train capacity, $N_0$ denotes the current number of passengers on this train, $N$ denotes the
number of passengers who get on the platform earlier than passenger $i$ at the same platform, and $x \in [0, 1]$ is a tuning parameter. We focus on the response surface $y = f(x, \mu)$, where the response $y$ is the mean travel time of passengers who tap in at the first station, $x \in [0, 1]$ is the tuning parameter in (19) at the first station, and $\mu \in \mathcal{P}(1, 2), \tau = \{ \mu \in \mathcal{P}(\mathbb{R}) : \text{support of } \mu \subset [1, 2], \sup_{x,y \in [1,2], x \neq y} |f(\mu(x) - F(\mu(y))/(|x-y| \leq \tau) \}$ represents the distribution of access time at the first station. The constant $\tau$ can be specified by experience on the upper bound of the density function of access time. Here we fix $\tau = 3$. Other inputs of the simulator are fixed. The simulation contains 72,000 passengers in the day.

We first use a 40-run maximin $W_{2,2}$-distance LH-type design which is the same as LH-L2 in Section 5 to compute the corresponding values of $y$. Due to the randomness of the simulation, each value of $y$ is obtained with 10,000 replicates. We then use SK and UK to build prediction models for the response. To compare the two models, we present the leave-one-out prediction values of $y$. The corresponding 90% prediction intervals are shown in Section E of the supplementary materials. It can be seen that UK is better than SK in most runs. SK and UK yield leave-one-out mean squared prediction errors $4.8 \times 10^{-3}$ and $7.3 \times 10^{-4}$, respectively.

Consequently, we adopt the response surface $\hat{f}$ constructed by UK based on all the data. Two profile curve of $\hat{f}(x, \mu)$ with fixed $\mu$ are shown in Figure 8. The limits of the 90% prediction intervals, which are computed by (18), are also shown in the figure. We can see that, as the tuning parameter $x$ increases, the travel times of the passengers from the first station have an increasing trend.

In the following study we can develop sensitivity analysis methods to quantify the influence of the two input parameters $x$ and $\mu$ on the response. Conventional methods such as the Sobol’ index method (Sobol’ 1993) cannot handle the distributional inputs. A feasible method is to use the idea proposed by Xiong et al. (2018) that computes the sensitivity index with permutations of the corresponding input observations. By this method we compute the sensitivity indices of $x$ and $\mu$. They are both small ($< 10^{-4}$), and this indicates that the two inputs are active. Another important issue is to calibrate the unobserved parameter $x$ with real data from the automatic fare collection system. We will study it when obtaining real data in the future.

7. Discussion

In this article we have proposed design and modeling methods for computer experiments with both numeral and distributional inputs. We use the Wasserstein distance to unify the mixed inputs, and then the proposed methods can be viewed as straightforward extensions of the conventional methods for the Euclidean space. This makes our methods easy to understand and to implement.

There are several further topics we can follow in the future. It may be interesting to extend lattice-based (He 2021) methods to construct space-filling designs in the probability measure space. Designs defined by other distribution distances or model-based designs can also be considered. Methods for sensitivity analysis, parameter calibration, and response optimization with both numeral and distributional inputs can be developed (Baker et al. 2022). In addition, possible directions include the study of computer simulations with distributional outputs and/or multidimensional distributional inputs based on the Wasserstein distance, which calls for strategies to overcome the difficulties in computation (Peyré and Cuturi 2019).

Supplementary Materials

Supl: This is a pdf file providing technical proofs and additional numerical results in the article.
DACEnd: This is a zip file containing Matlab codes to perform the design construction and Gaussian process modeling methods in the article.

Acknowledgments

We thank the editors and referees for constructive comments which lead to a significant improvement of this article.

Funding

Xiong’s work is partially supported by National Key R&D Program of China (grant nos. 2021YFA1000300, 2021YFA1000301, and 2021YFA1000303) and the National Natural Science Foundation of China (grant no. 12171462). Cui’s work is partially supported by the National Natural Science Foundation of China (grant nos. 12171234, 11790272), the Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), and the Fundamental Research Funds for the Central Universities.
ORCID

Shifeng Xiong  http://orcid.org/0000-0002-9636-038X

References

Arjovsky, M., Chintala, S., and Bottou, L. (2017), “Wasserstein Generative Adversarial Networks,” in Proceedings of the 34th International Conference on Machine Learning (Vol. 70), pp. 214–223. [408]

Bachoc, F., Gamboa, F., Loubes, J.-M., and Venet, N. (2018), “A Gaussian Process Regression Model for Distribution Inputs,” IEEE Transactions on Information Theory, 64, 6620–6637. [408,412]

Bachoc, F., Suvironkova, A., Ginsbourger, D., Loubes, J.-M., and Spokoiny, V. (2020), “Gaussian Processes with Multidimensional Distribution inputs via Optimal Transport and Hilbertian Embedding.” Electronic Journal of Statistics, 14, 2742–2772. [408,412]

Baker, E., Barbillón, P., Fadikar, A., Gramacy, R. B., Herbei, R., Higdon, D., Huang, J., Johnson, L. R., Ma, P., Mondal, A., Pires, B., Sacks, J., and Sokolov, V. (2022), “Analyzing Stochastic Computer Models: A Review with Opportunities,” Statistical Science, 37, 64–89. [406,416]

Betancourt, J., Bachoc, F., Kleina, T., Idier, D., Pedreros, R., Rohmer, J. (2020), “Gaussian Process Metamodeling of Functional-Input Code for Coastal Flood Hazard Assessment,” Reliability Engineering and System Safety, 198, 106870. [408]

Chen, J., Mak, S., Josep, V. R., and Zheng, C. (2021), “Function-on-Function Kriging, with Applications to Three-Dimensional Printing of Aortic Tissues,” Technometrics, 63, 384–395. [408]

De Boor, C. (1978), A Practical Guide to Splines, New York: Springer. [413]

Elefteriadou, L. (2014), An Introduction to Traffic Flow Theory, New York: Springer. [406]

Fang, K. T., Li, R. Z., and Sudjianto, A. (2005), Design and Modeling for Computer Experiments, London: Chapman Hall/CRC Press. [406]

Fang, K. T., Lin, D. K. J., Winker, P., and Zhang, Y. (2000), “Uniform Design: Theory and Application,” Technometrics, 42, 237–248 [410]

He, X. (2017), “Rotated Sphere Packing Designs,” Journal of the American Statistical Association, 112, 1612–1622. [410]

Hedayat, A., Sloane, N., and Stufken, J. (1999), Orthogonal Arrays, New York: Springer. [410]

Johnson, M., Moore, L., and Ylvisaker, D. (1990), Minimax and Maximin Distance Design, Journal of Statistical Planning and Inference, 26, 131–148. [409]

Joseph, V. R. (2016), “Space-Filling Designs for Computer Experiments: A Review,” Quality Engineering, 28, 28–35. [410]

Joseph, V. R., Gul, E., and Ba, S. (2015), “Maximum Projection Designs for Computer Experiments,” Biometrika, 102, 371–380. [410]

Loeppky, J. L., Sacks, J., and Welch, W. J. (2009), “Choosing the Sample Size of a Computer Experiment: A Practical Guide,” Technometrics, 51, 366–376. [413]

Li, C., Xiong, S., Sun, X., Qin, Y. (2022), “Bayesian Analysis for Metro Passenger Flows using Automated Data,” Mathematical Problems in Engineering, 2022, 992593. [415]

Mckay, M. D., Beckman, R. J., and Conover, W. J. (1979), “A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code,” Technometrics, 21, 239–245. [410]

Morris, M. D., and Mitchell, T. J. (1995), “Exploratory Designs for Computational Experiments,” Journal of Statistical Planning and Inference, 43, 381–402. [409]

Mu, W., and Xiong, S. (2017), “On Algorithmic Construction of Maximin Distance Designs,” Communications in Statistics: Simulation and Computation, 46, 7972–7985. [410]

Mu, W., and Xiong, S. (2018), “A Class of Space-Filling Designs and their Projection Properties,” Statistics & Probability Letters, 141, 129–134. [410]

Mo, B., Ma, Z., Koutsopoulos, H. N., and Zhao, J. (2021), “Calibrating Path Choices and Train Capacities for Urban Rail Transit Simulation Models using Smart Card and Train Movement Data,” Journal of Advanced Transportation, 2021, 5597130. [406]

Muehlenstaedt, T., Fruth, J., and Roustant, O. (2017), “Computer Experiments with Functional Inputs and Scalar Outputs by a Norm-based Approach,” Statistics and Computing, 27, 1083–1097. [408]

Nanty, S., Helbert, C., Marrel, A., Pêrot, N., and Prieur, C. (2016), “Sampling, Metamodeling, and Sensitivity Analysis of Numerical Simulators with Functional Stochastic Inputs,” SIAM/ASA Journal on Uncertainty Quantification, 4, 636–659. [406]

Owen, A. B. (1994), “Controlling Correlations in Latin Hypercube Samples,” Journal of the American Statistical Association, 89, 1517–1522. [410]

Peysé, G., and Cuturi, M. (2019), “Computational Optimal Transport,” Foundations and Trends in Machine Learning, 11, 355–607. [408,416]

Santner, T. J., Williams, B. J., and Notz, W. I. (2018), The Design and Analysis of Computer Experiments (2nd ed.), New York: Springer. [406,410,412,413]

Sobol’, I. M. (1993), “Sensitivity Analysis for Non-linear Mathematical Models,” Mathematical Modeling and Computational Experiment, 1, 407–414. [416]

Squazzoni, F., Jager, W., and Edmonds, B. (2014), Social Simulation in the Social Sciences: A Brief Overview, Social Science Computer Review, 32, 279–294. [406]

Tan, M. H. Y. (2019), “Gaussian Process Modeling of Finite Element Models with Functional Inputs,” SIAM/ASA Journal on Uncertainty Quantification, 7, 1133–1161. [408]

Tseng, P. (2001), “Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization,” Journal of Optimization Theory and Applications, 109, 475–494. [410]

Villani, C. (2009), Optimal Transport, Old and New, Berlin: Springer. [408]

Xiong, S. (2021), “The Reconstruction Approach: From Interpolation to Regression,” Technometrics, 63, 225–235. [413]

Xiong, S. He, X. He, Y., and Mu, W. (2018), “Sensitivity Analysis Using Permutations,” Statistica Sinica, 28, 817–837. [416]

Xiong, S., Li, C., Sun, X., Qin, Y., and Wu, C. F. J. (2022), “Statistical Estimation in Passenger-to-Train Assignment Models based on Automated Data,” Applied Stochastic Models in Business and Industry, 38, 287–307. [407,415]

Xiu, D. (2010), Numerical Methods for Stochastic Computations: A Spectral Method Approach, Princeton: Princeton University Press. [406]

Zhu, Y., Koutsopoulos, H. N., and Wilson, N. H. M. (2017), “A Probabilistic Passenger-to-Train Assignment Model based on Automated Data,” Transportation Research Part B: Methodological, 104, 522–542. [407]