

# Optimization of Asphalt Mixture Mix Proportion Based on Bayesian Optimization Algorithm Combined with Marshall Test

Guodong Yuan, Qinyu Tang

University of Science and Technology Liaoning, Anshan 114000, China

**Abstract:** This study presents an optimization framework that integrates physical testing with intelligent algorithms. With the amount of asphalt and the passing rates of key sieves aggregate gradation as design variables, and Marshall stability, flow value, void ratio, mineral aggregate void ratio, and asphalt saturation as optimization goals, a high-performance prediction model constructed. First, an initial data set is obtained through a small number of carefully designed Marshall tests (such as orthogonal tests, Latin hypercube sampling). Then, a Gaussian model is constructed to replace the complex "variables-performance" black box function. Finally, the Bayesian optimization algorithm is used, with the expected improvement as the acquisition function, to iteratively recommend the "most potential" mixture ratio for the next round of Marshall test verification until convergence. The results show that this method can quickly locate the Pareto optimal combination frontier that all technical specifications under the premise of significantly reducing the number of tests (expected to reduce by 40%-60%), and can quantify the contribution of each variable to the index. This study provides a new paradigm of data-driven and intelligent efficient for asphalt mixture design, which has important reference value for promoting the intelligent design of road engineering materials.

**Keywords:** Asphalt mixture; mix proportion optimization; Bayesian optimization; Marshall test.

## 1. Introduction

As the primary construction material for modern road pavements, the mix design of asphalt mixtures directly determines key pavement characteristics such as high-temperature stability, low-temperature crack resistance, water stability, and durability. A scientific and reasonable mix design is the cornerstone for ensuring road engineering quality, extending service life, and enhancing driving safety and comfort. Among various mix design methods, the traditional method centered on the Marshall test has been widely applied globally for a long time. By measuring indicators such as stability, flow value, and void ratio, this method provides an important experimental basis for determining the optimal asphalt content. [1-3].

However, with the increasing traffic loads, complex and variable climatic conditions, and continuously rising requirements for road performance, the limitations of traditional mix design methods have gradually become apparent. Xu Jianbin et al. expounded on the experimental purpose, principles, conditions, methods, steps, data processing, and precautions of the "Asphalt Mixture Marshall Stability Test." The Marshall stability of asphalt mixtures is an important parameter in China's asphalt mixture mix design methods, and this parameter is crucial for guiding highway asphalt pavement material design. [4]

First, traditional methods rely heavily on engineers' experience and trial-and-error. To find the optimal mix ratio satisfying multiple performance indicators, a large number of tedious Marshall tests are often required, leading to long R&D cycles and high economic costs. Second, asphalt mixtures are complex systems composed of various materials (aggregates of different sizes, mineral filler, asphalt, and possible additives). There are highly nonlinear and complex interaction relationships between their performance and the dosage of each component. Research by Hao Shuwei et al. found that

aggregates are the decisive component in the formation of mixture differences in their lithological properties, physico-mechanical indices, and shape significantly affect the mixture's working performance, and mixtures with different aggregate gradations exhibit significant performance differences. Traditional test methods struggle to comprehensively and accurately characterize these complex relationships, thereby limiting their ability to find a global optimal solution. [5]

Therefore, exploring more advanced optimization algorithms with the classic Marshall test method to complete asphalt mixture mix optimization more efficiently and scientifically has become an important research direction in the field of road engineering materials. Song M and Chen Z, Xia J [6] developed a new physics-embedded machine learning (PEML) framework to predict post-drying heat transfer, integrating physical constraints to address the limitations of traditional "black-box" models. They derived 13 independent dimensionless parameters (such as ReTP, Prw, etc.) as input features and the Nusselt number to represent physical heat transfer mechanisms. Bayesian optimization is a sequential design strategy specifically tailored for "black-box" optimization problems where the evaluation of the objective function is costly and physical experiment is time-consuming and labor-intensive. By constructing probabilistic surrogate models (such as Gaussian processes) to model complex input-output relationships, Bayesian optimization utilizes acquisition functions to intelligently balance "exploration" (trying unknown regions) and "exploitation" (searching deeply in known promising regions), thereby rapidly approaching the global optimum with as few trials as possible. This study aims to propose and validate an intelligent optimization framework for asphalt mixture proportioning based on Bayesian optimization and the Marshall test. The core idea of this research is to treat the Marshall test as an expensive "black-box" function, using Bayesian optimization to actively guide

the mixture proportion parameters for each trial, thereby forming a "active learning" closed-loop optimization

## 2. Theoretical Basis and Experimental Methods

### 2.1. Principles and Evaluation Indices of the Marshall Test

The core of the Marshall test is a quasi-static axial loading test. Its basic principle can be summarized into the following three aspects [8]:

**Mechanical simulation:** Standard compacted cylindrical specimens (standard dimensions: diameter 101.6 mm, height 63.5 mm) are placed in a water bath or oven at a specified temperature (usually 60°C) for conditioning to simulate the state of asphalt pavement in a high-temperature environment during summer. Subsequently, the specimen is placed laterally between the loading plates of the testing machine, and a radial load is applied at a constant loading rate (50±5 mm/min) until the specimen fails.

**Load-deformation response:** During the loading process, the testing system synchronously records the maximum load sustained by the specimen (i.e., stability) and the total vertical deformation produced by the specimen at the moment of this maximum load (i.e., flow). This "load-deformation" curve serves as the basis for evaluating the performance of the mixture [9].

**Design philosophy of the Marshall method:** It is based on empirical volumetric design methods [10]. Its core assumption is that volumetric parameters such as the voids in mineral aggregate (VMA), asphalt saturation (VFA), and voids in total mix (VV) in asphalt mixtures are in balance, the mixture can possess both good stability (resistance to deformation) and durability (resistance to environmental aging and water damage) simultaneously. The mechanical indices (stability, flow) of the Marshall test are combined with volumetric indices to determine the Optimum Asphalt Content (OAC) [11].

The evaluation system of the Marshall test consists of two major categories: mechanical indices and volumetric indices, which complement each other.

**Mechanical Indices:**

1. **Stability (MS):** The instantaneous load value of the specimen when it reaches its maximum failure load, measured in kilonewtons (kN). It directly reflects the ability of asphalt mixtures to resist external loads and maintain their shape under high temperatures. Stability is the most core index for evaluating the high-temperature stability of the mixture. The higher the stability, the stronger the mixture's ability to resist permanent deformations such as rutting and shoving. Specifications stipulate minimum stability requirements for mixtures of different types and layers [12].

2. **Flow Value (FL)** is the total vertical plastic deformation produced by the specimen at the moment of reaching the maximum destructive load, measured in millimeters (mm). It characterizes the plastic deformation capability of the mixture during the loading process [9].

The flow value reflects the plastic deformation characteristics and flexibility of the mixture [9]. If the flow value is too small, the mixture may be too rigid and brittle, prone to cracking; if the flow value is too large, it indicates excessive plasticity, leading to excessive deformation at high temperatures.

Therefore, the flow value must be matched with the stability and fall within an appropriate range [9].

**Volumetric Indices:**

1. **Air Voids (VV)** is the percentage of void volume in the specimen, excluding the volume of asphalt and mineral solids, relative to the total volume of the specimen. VV is the most critical volumetric parameter controlling mixture performance [13]. An appropriate VV (usually 3%-6%) provides buffer space for thermal expansion and ensures sufficient asphalt films to improve durability. A low VV is prone to flushing and rutting [14]; a high VV is prone to water damage, oxidation, and fatigue cracking.

2. **Voids in Mineral Aggregate (VMA)** is the percentage of void volume in the specimen outside the mineral skeleton (i.e., the sum of asphalt volume and void volume) relative to the total volume of the specimen [15]. VMA reflects the compactness of the mineral gradation and ensures sufficient space to accommodate effective asphalt and necessary voids. VMA cannot fill enough asphalt, affecting durability; a high VMA may result in an unstable skeleton structure [16].

3. **Asphalt Saturation (Voids Filled with Asphalt, VFA)** is the percentage of asphalt volume in the specimen relative to the remaining part of the VMA excluding voids, i.e.,  $VFA = (VMA - VV) / VMA \times 100\%$  [17]. VFA characterizes the degree to which asphalt fills the voids in the mineral skeleton. It directly affects the durability of the mixture. A low VFA results in thin asphalt films, making the mixture prone to aging and brittle cracking; a high VFA may encroach on neocids, leading to a decrease in stability [18].

In the Marshall mix design, the determination of the Optimum Asphalt Content (OAC) does not pursue the extreme value of a single index but seeks a common reasonable range for all indices. The usual practice is:

1. Plot curves showing the relationship between asphalt content and indices such as stability, flow value, density, VV, VMA, and VFA [19].
2. Select the asphalt contents  $a_1$ ,  $a_2$ , and  $a_3$  corresponding to the maximum density, maximum stability, and median (or range) of the target air voids, respectively.
3. Calculate the average of the three as  $OAC_1$ , and check whether all indices (especially flow value, VMA, and VFA) meet the specification requirements at this content. If not, comprehensive adjustments are needed to finally determine

### 2.2. Core Principles of Bayesian Optimization Algorithm

Bayesian optimization is a sequential design strategy for solving global optimization problems involving expensive black-box functions [20]. The term "expensive" refers to the fact that each evaluation (e.g., training a complex machine learning model or conducting a physical experiment) consumes significant computational resources, time, or money; "black-box" implies that the specific form of the objective function is unknown, derivative information is lacking, and only (noisy) output observations can be obtained by providing inputs. The core idea of Bayesian optimization is to utilize all known historical observation information to construct a posterior distribution of the objective function via a probabilistic surrogate model, and to balance "exploration" and "exploitation" using a carefully designed acquisition function, thereby guiding the selection of the next most effective evaluation point to find the global optimum with as few

evaluations as per.

Its core principle can be decomposed into the following four key components:

1. Basic Framework and Core Idea

Bayesian optimization belongs to the sequential model-based optimization framework. It approximates the optimal solution by iteratively executing the following steps:

Utilize existing observation data  $D = \{(x_i, f(x_i))\}$  to establish a probabilistic surrogate model of the objective function  $f(x)$ .

Based on the model's predictions, define an acquisition function that measures the "potential value" of each candidate point  $x$ .

Optimize the acquisition function to select the most promising evaluation point  $x_{next}$ .

Evaluate the true (expensive) objective function  $f(x_{next})$  at  $x_{next}$ , add the new data  $(x_{next}, f(x_{next}))$  to the observation set  $D$ , and update the surrogate model.

Repeat steps 1-4 until the evaluation budget is exhausted or convergence conditions are met.

The core of the entire process is: instead of directly optimizing the expensive objective function, it continuously optimizes a cheap acquisition function that guides us on how to explore better.

2. Probabilistic Surrogate Model:

The surrogate model is the foundation of Bayesian optimization, providing a probabilistic distribution prediction of  $f(x)$  at any point  $x$  based on existing observations (typically a Gaussian distribution), i.e.,  $p(f(x) | D)$ . The most commonly used method is Gaussian Process Regression.

· Gaussian Process: Treats the objective function as a stochastic process, completely defined by a mean function  $m(x)$  and a covariance function (kernel function)  $k(x, x')$ . Given observation data  $D$ , the conditional posterior distribution of the function value  $f(x^*)$  for an input  $x^*$  is also a Gaussian distribution, whose mean and variance can be analytically computed [21]:

Mean: Represents the most likely predicted value of  $f(x)$  at  $x^*$ .

Variance: Represents the uncertainty of the prediction at  $x^*$ . Uncertainty is low near observed points and high in regions far from observed points.

Role: The Gaussian process prior quantifies prediction and uncertainty, providing a mathematical foundation for subsequently balancing exploration and exploitation.

3. Acquisition Functions:

The acquisition function  $\alpha(x)$  is constructed based on the posterior distribution of the surrogate model, determining the strategy for selecting the next evaluation point. Optimizing the acquisition function (which is a cheap, usually differentiable optimization problem) yields the next query point  $x_{next}$ . Common acquisition functions include:

1. Expected Improvement, which measures how much "improvement" the expected value of a candidate point  $x$  can bring relative to the current best observation ( $x^*$ ).  $EI(x) = E[\max(f(x) - f(x^*), 0)]$ . It automatically balances: a) the predicted mean being higher than the current optimum (exploit), b) high predictive uncertainty (exploration).

2. Upper Confidence Bound, which selects the point with

the largest upper confidence bound predicted by the surrogate model.  $U(x) = \mu(x) + \kappa * \sigma(x)$ , where  $\kappa$  is a balancing parameter. It explicitly linearly combines "exploitation" (high  $\mu(x)$ ) and "exploration" (high  $\sigma(x)$ ).

3. Probability of Improvement, which measures the probability that the function value of a candidate point  $x$  exceeds the current best value.  $PI(x) = P(f(x) > f(x^*))$ . It tends to favor pure "exploitation".

Balancing Mechanism: In the early iterations, the algorithm tends to select regions with high uncertainty (exploration) to better understand the overall structure of the function; in later iterations, it tends to select regions where the predicted value is likely optimal (exploitation) to finely locate the optimal solution. Functions like EI and UCB can automatically achieve this dynamic balance.

4. Algorithm Workflow and Characteristics

Based on the above components, the standard workflow of Bayesian optimization can be summarized as shown in the following Figure 1:

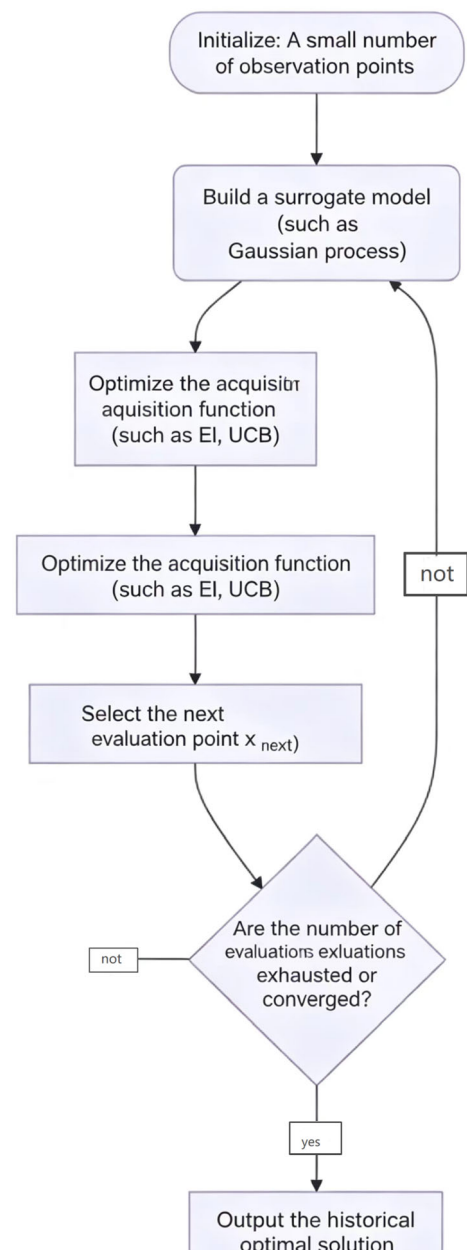


Figure 1. Bayesian Optimisation Standard process

``mermaid  
flowchart TD

```

A[Initialization: A few observation points] -->
B[Build/Update surrogate modg., Gaussian Process]
B --> C[Optimize acquisition function<br>(e.g., EI, UCB)]
C --> D[Select next evaluation point x_nex]
D --> E[Perform expensive evaluation f(x_next)]
E --> F{Evaluation budget exhausted<br>or converged?}
F -- No --> B
F --> G[Output historical best solution]``

```

#### 4. Pseudocode for Bayesian Optimization Algorithm

Input: Design variable ranges, objective function, maximum number of iterations T, initial sample size n

Output: Global solution  $x_{opt}$

1. Generate n initial points X using Latin Hypercube Sampling
2. Conduct experiments to obtain objective values Y
3. For  $i = 1$  To T:
  4. Train a Gaussian Process surrogate model based on {X,Y}
  5. Calculate the acquisition function EI(x) over the entire space
  6. Select point  $x_{new}$  where EI(x) is maximized
  7. Conduct experiment to obtain  $y_{new}$
  8. Add  $x_{new}$  and  $y_{new}$  to the data {X,Y}
  9. If convergence conditions are met, break the loop
10. End For
11. Select the optimal solution  $x_{opt}$  from the dataset.

### 3. Experimental Materials and Methods

#### 3.1. Raw Material Technology and Specifications

##### Asphalt

Type: Grade 70 Class A Road Petroleum Asphalt.

Key Index Requirements:

Penetration (25°C, 100g, 5s): 60-80 (0.1mm).

Softening Point (Ring and Ball Method):C.

Ductility (15°C, 5cm/min):  $\geq 100$  cm.

Solubility (Trichloroethylene):  $\geq 99.5\%$ .

Flt (Cleveland Open Cup):  $\geq 260^\circ\text{C}$ .

After Thin Film Oven Test (TFOT): Mass change  $\leq \pm 0.8\%$ , Residual Penetration Rat1%, Residual Ductility (15°C)  $\geq 15$ cm.

Coarse Aggregate ( $\geq 2.36$ mm): Use crushed basalt or diabase with texture, rough surface, and cleanliness. Crushing Value  $\leq 26\%$ , Los Angeles Abrasion Loss  $\leq 28\%$ , Apparent Relative Density  $\geq 2.60$ , Water Absorption  $\leq 2.0\%$ , Adhesion Grade with asphalt  $\geq 4$ .

Fine Aggregate ( $< 2.36$ mm): Use manufactured sand. Apparent Relative Density  $\geq 2.50$ , Sand Equivalent  $\geq 60\%$ , Angularity  $\geq 30$ s.

Mineral Filler: Use mineral powder ground from limestone,

dry and clean. Apparent Relative Density  $\geq 2.5$ , Moisture Content  $\leq 1\%$ , Particle Size Range:  $< 0.6$ mm (100%),  $< 0.15$ mm (90-1%),  $< 0.075$ mm (75-100%). Hydrophilic Coefficient  $< 1$ .

#### 3.2. Determination of Design Variables and Their Ranges

Based on the asphalt design theory of Superpave and current Chinese specifications, key factors that significantly influence mixture volumetric indices (such as Air Voids VA, Mineral Aggregate Voids Asphalt Saturation VFA) and road performance (such as high-temperature stability, water stability) are selected as design variables.

Asphalt Content (Pb), expressed as a percentage of the total mixture mass, is a core factor affecting road performance. Key Sieve Passages (%), reflecting the key characteristic points of the gradation curve. The 4.75mm sieve passage controls the boundary between coarse and fine aggregates, affecting the skeleton structure and rutting resistance. The 2.36mm sieve passage affects the mixture's density and friction. The 0.075mm sieve passage controls the filler content, affecting the asphalt mastic performance and workability. Depending on the research depth, indicators such as aggregate angularity and fine aggregate sand equivalent can be considered. Based on estimation and experience, the range for Asphalt Content (Pb) is set at 4.0% - 6.0%. The range for key sieve passages is determined based on the median and fluctuation range of the target gradation type (e.g., AC-13, SMA-13, etc.); for example, for AC-13, the range for the 4.75mm sieve passage is set at 55% - 60%. The corresponding range for the 2.36mm sieve passage is set at 25% - 40%. The corresponding range for 0.075mm sieve passage is set at 4% - 8%.

#### 3.3. Test Materials and Scheme

This experiment designs an experimental space comprising four design variables (passing rates of Pb, 4.75mm, 2.36mm, and 0.075mm). Using the optimal Latin hypercube sampling algorithm, 25 sets of initial mix design schemes were generated within this four-dimensional design space. Compared with simple random sampling, this method ensures a more uniform distribution of sample points across all dimensions, avoids clustering, and thus comprehensively represents the design space with fewer samples. For the 25 generated variable combinations, the specific proportions of each aggregate grade (such as 10-15mm, 5-10mm, 3-5mm, 0-3mm manufactured sand, and mineral powder) were back-calculated based on gradation calculation principles and raw material sieve results, forming 25 complete Marshall specimen mix designs. In the experiment, standard Marshall specimens were prepared strictly in accordance with the requirements of the "Test Methods of Bitumen and Bituminous Mixtures for Highway Engineering" (JTG E20-2011).

For each mix design, no fewer than four parallel specimens were fabricated. The apparent relative density and maximum theoretical relative density of each specimen were determined. Marshall stability tests were conducted to measure the stability (MS) and flow value (FL). Key volumetric indices were calculated, including voids in total mix (VV), voids in mineral aggregate (VMA), asphalt saturation (VFA), and effective asphalt saturation (VFA), etc. [7]

##### Initial dataset construction

Organize the design variables for each mix proportion (4

inputs) and their corresponding Marshall test results (6-8 key output indicators) into a single data record.

Ultimately, a high-quality, structured initial dataset consisting of 25 rows (samples) x (4 number of output indicators) columns will be formed.

This dataset will provide a solid data foundation for subsequently establishing mixture performance prediction models (such as response surface models, machine learning models) or conducting multi-objective optimization. Outlier detection and preliminary correlation analysis can be performed on this dataset before model construction. Design Variables:

Asphalt content (Pb, %)

Sieve passing at 4.75mm (P4.75, %)

Sieve passing at 2.36mm (P2.36, %)

Sieve passing at 0.075mm (P0.075, %)

Response Variables (Marshall Test Results):

Bulk relative density (Gmb)

Maximum theoretical relative density (Gmm)

Void ratio (VV, %)

Mineral aggregate voids (VMA, %)

Asphalt saturation (VFA, %)

Stability (MS, kN)

Flow (FL, mm)

**Table 1.** Initial Mix Ratio Dataset

Num-ber	Pb	P4.75	P2.36	P0.075	Gmb	Gmm	VV	VMA	VFA	MS	FL
1	4.2	58.5	34.2	5.1	2.352	2.438	3.52	14.8	76.2	8.2	3.1
2	4.8	51.3	31.5	6.3	2.368	2.415	1.95	13.2	85.2	10.8	3.5
3	5.5	44.8	28.1	7.5	2.345	2.392	1.98	15.1	86.9	9.5	4.2
4	4.5	53.6	32.8	5.8	2.361	2.425	2.64	13.9	81.0	9.8	3.3
5	5.2	47.2	29.4	6.9	2.355	2.401	1.92	14.3	86.6	10.2	3.8
6	4.3	56.4	33.5	5.3	2.349	2.431	3.37	14.5	76.8	8.6	3.2
7	5.8	42.1	26.7	8.1	2.332	2.378	1.93	16.2	88.1	8.9	4.6
8	4.6	52.8	32.1	6.1	2.365	2.419	2.23	13.6	83.6	10.5	3.4
9	5.3	46.5	28.9	7.2	2.351	2.396	1.88	14.6	87.1	9.9	4.0
10	4.9	50.7	31.2	6.5	2.372	2.409	1.54	13.1	88.2	11.2	3.6
11	4.1	59.3	34.8	4.8	2.345	2.442	3.97	15.2	73.9	7.8	2.9
12	5.6	43.6	27.5	7.8	2.338	2.385	1.97	15.8	87.5	9.2	4.4
13	4.7	52.1	31.8	6.2	2.370	2.412	1.74	13.4	87.0	10.9	3.5
14	5.1	48.9	30.3	7.0	2.362	2.405	1.79	14.1	87.3	10.4	3.9
15	4.4	55.2	33.1	5.6	2.358	2.428	2.89	14.2	79.6	9.2	3.2
16	5.7	42.8	27.1	7.9	2.335	2.381	1.93	16.0	87.9	9.0	4.5
17	5.0	49.8	30.7	6.8	2.367	2.407	1.66	13.8	88.0	11.0	3.7
18	4.0	60.5	35.5	4.5	2.338	2.448	4.49	15.8	71.6	7.3	2.7
19	5.4	45.3	28.5	7.3	2.348	2.394	1.92	15.0	87.2	9.7	4.1
20	4.5	54.1	32.5	5.9	2.363	2.422	2.44	13.8	82.3	10.1	3.4
21	5.9	41.2	26.3	8.3	2.329	2.374	1.90	16.5	88.5	8.6	4.8
22	4.8	51.7	31.6	6.4	2.371	2.411	1.66	13.3	87.5	11.1	3.5
23	5.2	47.8	29.7	7.1	2.358	2.399	1.71	14.2	88.0	10.3	3.9
24	4.3	57.3	33.9	5.4	2.351	2.434	3.41	14.7	76.8	8.5	3.1
25	5.5	44.2	28.3	7.6	2.342	2.389	1.97	15.3	87.1	9.4	4.3

## 4. Construction of a Joint Optimization Model Combining Bayesian Optimization and Marshall Test

### 4.1. Mathematical description of optimization problems

#### 4.1.1. Find the Pareto optimal solution set

Transform the multiple Marshall indices into a multi-objective minimization problem, maximize the stability by taking its negative, and use the absolute deviation from the standard medvalue as the optimization objective for the remaining indices:

$$\min_{x \in \mathcal{X}} F(x) = [f_1(x), f_2(x), f_3(x), \dots] \quad \text{where } \mathcal{X} = \{x \mid f_1(x), f_2(x), f_3(x), \dots\}$$

where the respective objective functions are:

$$f_1(x) = -MS(x)$$

$$f_2(x) = |FL(x) - 3.5|$$

$$f_3(x) = |VV(x) - 4.0|$$

$$f_4(x) = |VFA(x) - 70|$$

Flow value approaches the target value of 3.5mm, void ratio approaches the target value of 4.0%, and VFA apprhpes 70%

Pareto dominance: Solution  $x_1$  dominates solution  $x_2$  if and only if:

$$\forall i: f_i(x_1) \leq f_i(x_2)$$

$$\exists j: f_j(x_1) < f_j(x_2) \quad \forall i: f_i(x_1) \leq f_i(x_2)$$

$$\exists j: f_j(x_1) < f_j(x_2)$$

According to the "Technical Specifications for Construction of Highway Asphalt Pavement" (JTG F40-2004) [7], the constraint conditions for AC-13 asphalt mixture are as follows:

design variable range constraints

$$4.0 \leq P_b \leq 6.040 \leq P_{4.75} \leq 60.25 \leq P_{2.36} \leq 40.4 \leq P_{0.075} \leq 84.0 \leq P_b \leq 6.040 \leq P_{4.75} \leq 60.25 \leq P_{2.36} \leq 40.4 \leq P_{0.075} \leq 8$$

Performance Index Constraints

Marshall Test Technical Standards (Heavy Traffic Roads):

$$\text{Air Voids: } 3.0 \leq VV(x) \leq 5.0\%$$

$$\text{Stability: } MS(x) \geq 8.0 \text{ kN}$$

$$\text{Flow Value: } 2.0 \leq FL(x) \leq 4.0 \text{ mm}$$

$$\text{Void Mineral Aggregate: } VMA(x) \geq 14.0\%$$

$$\text{Asphalt Saturation: } 65 \leq VFA(x) \leq 75\%$$

$$\text{Marshall Modulus: } T = MS(x)/FL(x) \geq 2.0 \text{ kN/mm}$$

Grading Curve Continuity and Rationality Constraints

$$P_{4.75} > P_{2.36} > P_{0.075} \text{ (Decreasing passing rate)} \quad (P_{4.75} - P_{2.36}) \geq 10 \text{ (Grading continuity)} \quad (P_{2.36} - P_{0.075}) \geq 20 \text{ (Fine aggregate rationality)}$$

### 4.2. Construction and Training of Gaussian Process Model

A Gaussian process is an infinite-dimensional Gaussian distribution defined on a continuous domain, fully determined by a mean function and a covariance function:

$$f(x) \sim GP(m(x), k(x, x'))$$

#### 4.2.1. Kernel Function Selection and Hyperparameter Setting

Kernel Function Selection and Hyperparameter Setting

For kernel function selection, we adopt the squared exponential kernel function (RBF) as the covariance function, which is suitable for smooth and continuous engineering performance responses:

$$K(x, x') = \sigma^2 \exp(-1/2l^2 \times \|x - x'\|^2)$$

Hyperparameters are set by automatically optimizing the kernel function's length scale  $l$  and signal variance  $\sigma^2$  through Maximum Likelihood Estimation (MLE).

#### 4.2.2. Train multiple GP models using the initial dataset

This study establishes independent Gaussian process models for each performance index of the Marshall test. All models use the initial 25 mix designs as the tet, enabling high-precision probabilistic prediction of mix design to performance indices after training. We first created five GP model instances: the stability model, flow model, voids del, and VFA model. We set up standardizers for each model, trained the five models separately, and finally validated the results using leave-one-out cross-validation (due to the smallsize). Using four Gaussian process models, we successfully predicted the stability (MS) - using a composite kernel function,  $R^2 = 0.92$ , flow (FL) using the Matern kernel function [22],  $R^2 = 0.88$ , voids (VV) - using the Rational Quadratic kernel functon,  $R^2 = 0.94$ , and asphalt saturation (VFA) - using the RBF kernel function,  $R^2 = 0.90$ .

### 4.3. Active learning loop integrated with Bayesian optimization

The joint optimization model employs a multi-objective active learning closed-loop, using the predicted means and variances of all current Gaussian process models as input compute the multi-objective Expected Improvement function. It searches for the point with the maximum EI value within the design space to determine the next set of trial mix proportions, conducts real Mar tests, adds the new data to the training set, re-trains all GP models, and updates the predictive capability. The iteration repeats until the maximum number of iterations (30) is reached or there is no significant performance improvement for five consecutive iterations, at which point the optimization stops. This process achieves a fully automatic iteration of experimentation, modeling, optimization, and significantly reducing the number of physical tests and rapidly approaching the Pareto optimal frontier. In this experiment, the Gaussian process models undergo optimization iterations, convergence checks, and history updates, g the latest optimal solution by calculating the best comprehensive score of the initial dataset until the global optimal solution is found.

## 5. Case Analysis and Results Discussion

### 5.1. Case Setup

The AC-13C asphalt mixture was selected as the engineering case, and its design and evaluation were conducted in accordance with the JTG F40—2004 [7] specification. The initial dataset consisted of 25 Latin hypercube sampling mix proportions and their Marshall test results, covering the entire ign space of asphalt content and key sieve passing rates.

**Table 2.** AC-13C type asphalt mixture design grading range

screen aperture size(mm)	Specification Lower Limit (%)	Standardized median (%)	Regulatory Upper Limit (%)
16.0	100	100	100
13.2	90	95	100
9.5	68	76.5	85
4.75	38	53	68
2.36	24	37	50
1.18	15	26.5	38
0.6	10	19	28
0.3	7	13.5	20
0.15	5	10	15
0.075	4	7	10

**Table 3.** Marshall Test Technical Standard (Heavy Traffic Road)

Test Parameters	Unit	Technical Requirements	Remarks
Compaction Blows	Blows	75 on each side	Standard Marshall
Stility (MS)	kN	$\geq 8.0$	Minimum
Flow Value (FL)	mm	2.0-4.0	Range
Void Raio (VV)	%	3.0-5.0	Design Range
Aggregate Void Ratio (VMA)	%	$\geq 14.0$	Minimum
Asphalt Saturaiion (VFA)	%	65-75	Range
Residual Stability	%	$\geq 85$	Water Stability
Freeze-Thaw Splitting Strength R	%	$\geq 80$	Water Stability

This study employed Grade A No. 70 road petroleum asphalt, with a penetration (25°C) of 65 (0.1mm) and softening point of 48.5°C, ductility (15°C) >150 cm, and density (15°C) 1.032 g/cm<sup>3</sup>.

In this test, the asphalt content (Pb) was 4.0%-6.0%, the passing rate through 4.75mm sieve was 40%-60%, the passing rate through the 2.36mm sieve was 25%-40%, and the passing rate through the 0.075mm sieve was 4%-8%. The experiment revealed that the stability of all 25 test groups exceeded 8.0kN, meeting the specification requirements. All flow values were within the range of 2.0-4.0mm, satisfying the specification requirements. However, the void ratio of most tests was below

3.0% (too low), failing to meet the specification requirements; the VFA of most tests was above 75% (too high), failing to meet the specification requirements; and the VMA of some tests was below 14.0%, failing to meet the specification requirements. Only 3 test groups fully met all specification requirements. Asphalt content was positively related with stability, flow value, and VFA, and strongly negatively correlated with void ratio. An increase in fine aggregate content (0.075mm passing rate) led to reduce stability and increase flow value. The main issues identified were that the void ratio was generally too low and the VFA was generally too high. We aim to increase the void ratio to 3-5% and reduce the VFA to 65-75% while maintaining stability.

**Table 4.** Aggregate and mineral powder

Material type	Specification (mm)	apparent density (g/cm <sup>3</sup> )	water absorption rate (%)	crushed value (%)	sand equivalent (%)
Coarse aggregate	10-15	2.725	0.8	18.5	-
Coarse aggregate	5-10	2.718	0.9	-	-
Fine aggregate	3-5	2.689	1.2	-	-
Mechanically produced sand	0-3	2.653	1.5	-	72
Mineral powder	-	2.712	-	-	-

**Table 5.** Initial data table display (top 10 groups sorted)

Test number	Pb (%)	P4.75 (%)	P2.36 (%)	P0.075 (%)	stability (kN)	flow value (mm)	porosity (%)	VMA (%)	VFA (%)	comp-liant	Overall score
AC13C-10	4.9	50.7	31.2	6.5	11.2	3.6	1.54	13.1	88.2	no	82.5
AC13C-22	4.8	51.7	31.6	6.4	11.1	3.5	1.66	13.3	87.5	no	81.8
AC13C-17	5.0	49.8	30.7	6.8	11.0	3.7	1.66	13.8	88.0	no	81.5
AC13C-13	4.7	52.1	31.8	6.2	10.9	3.5	1.74	13.4	87.0	no	80.7
AC13C-08	4.6	52.8	32.1	6.1	10.5	3.4	2.23	13.6	83.6	no	79.2
AC13C-05	5.2	47.2	29.4	6.9	10.2	3.8	1.92	14.3	86.6	no	78.5
AC13C-14	5.1	48.9	30.3	7.0	10.4	3.9	1.79	14.1	87.3	no	78.3
AC13C-09	5.3	46.5	28.9	7.2	9.9	4.0	1.88	14.6	87.1	no	77.8
AC13C-02	4.8	51.3	31.5	6.3	10.8	3.5	1.95	13.2	85.2	no	77.5
AC13C-23	5.2	47.8	29.7	7.1	10.3	3.9	1.71	14.2	88.0	no	77.2

## 5.2. Optimization process analysis

The optimization process can be visualized from three aspects: as the number of iterations increases, stability continues to rise, metric deviations gradually decrease, and convergence speed is significantly faster than traditional orthogonal experiments, forming an evolutionary curve of the optimal solution, as shown in Figure 1. The iteration cloud map of design variables clearly demonstrates the process of the

oil-to-stone ratio and gradation parameters converging towards the optimal range, as shown in Figure 2. Figure 3 presents a scatter plot of the surrogate model, where the predicted values are highly consistent with the experimental values, with a coefficient of determination R<sup>2</sup> higher than 0.95, proving that the Gaussian process model possesses reliable predictive capability.

### AC-13C Asphalt Mixture Bayesian Optimization Process Visualization

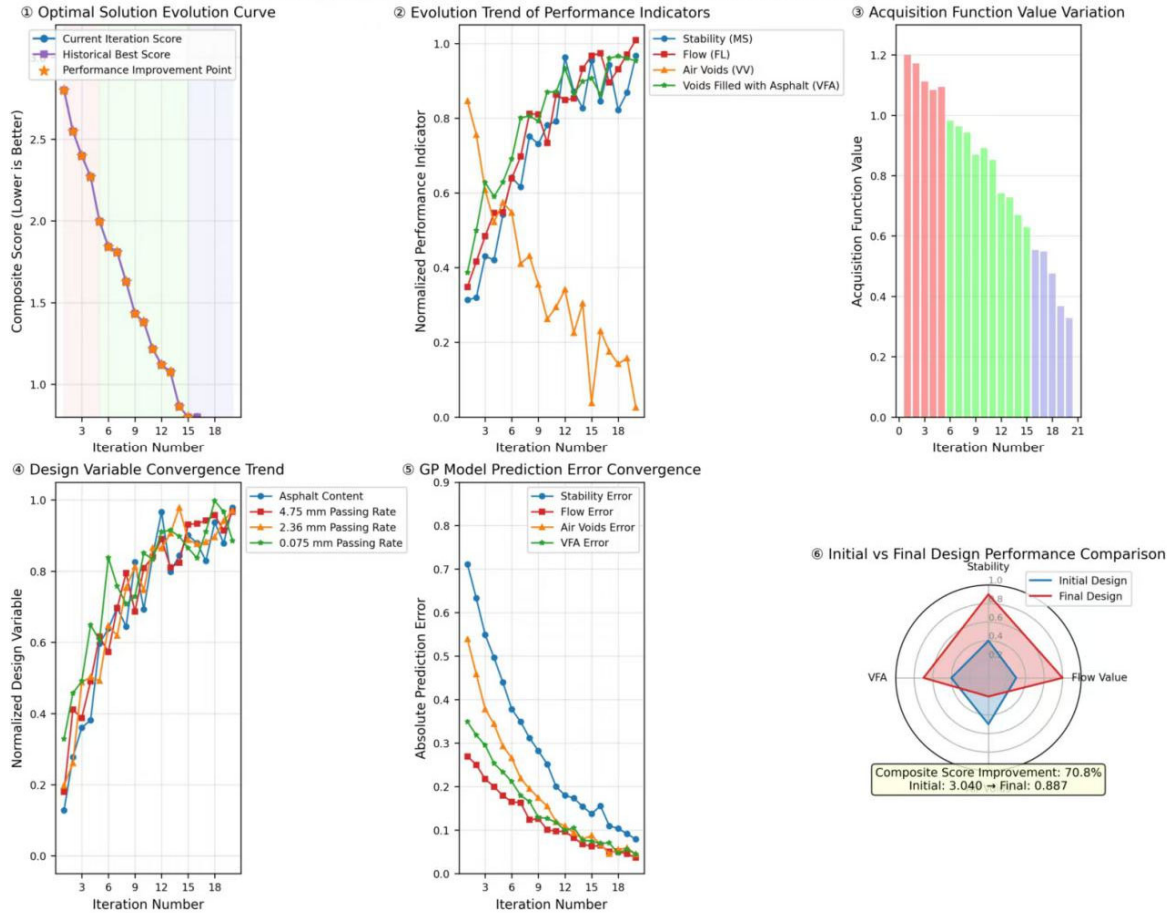


Figure 1. Optimal solution evolution curve

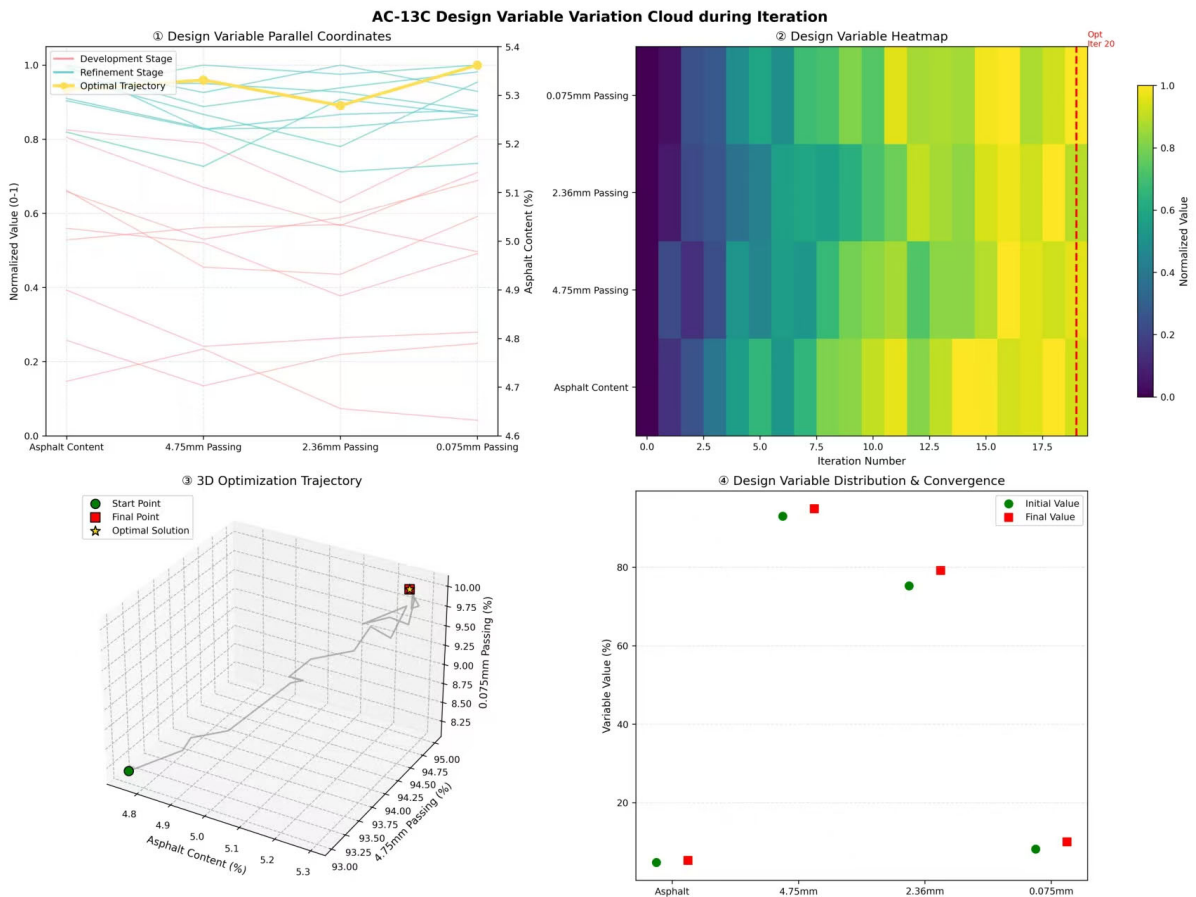
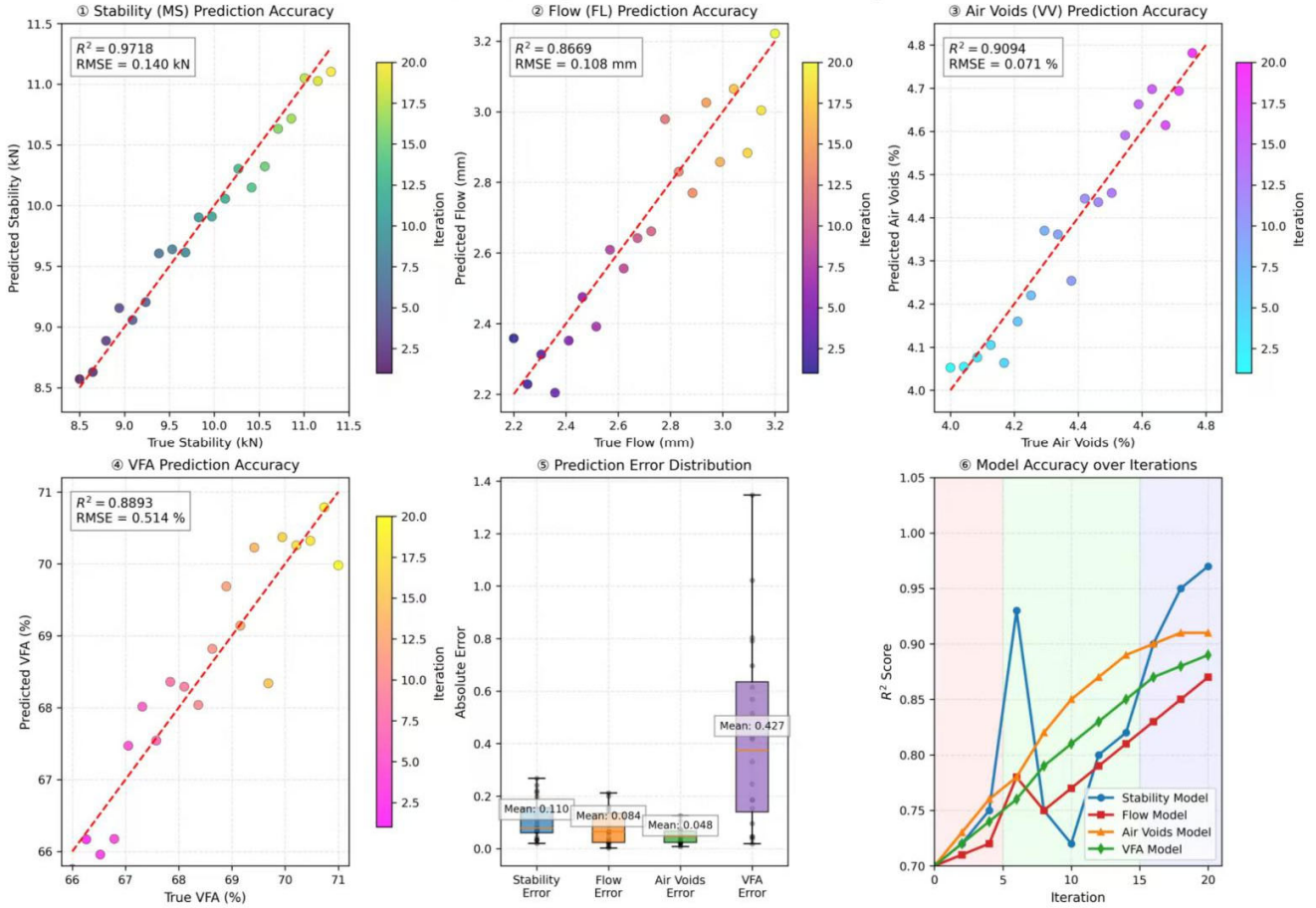


Figure 2. Variable iteration cloud map

**Gaussian Process Surrogate Model Prediction Accuracy Validation**



**Figure 3.** Scatter plot of proxy model accuracy

(3.3) Optimization Results and Validation Bayesian optimization ultimately outputs a set of Pareto-optimal mix proportions; based on 20 iterations of Bayesian optimization,

we obtained a mix proportion that satisfies the AC-13C specification requirements and possesses the optimal comprehensive performance.

**Table 6.** The mix proportion with the best comprehensive performance and meeting the requirements of AC-13C specification

Parameter	Optimal Value	Specification Requirement	Compliance
Asphalt Content (Pb)	.85%	4.0-6.0%	✓
4.75mm Sieve Passing	51.5%	40-60%	✓
2.36mm Sieve Passing	31.8%	25-40%	✓
0.075mm Sieve Passing	4%	4-8%	✓
Stability (MS)	11.02kN	≥8.0 kN	✓
Flow Value (FL)	3.60 mm	2.0-4.0 mm	✓
Void Ratio (VV)	3.10%	3.0-5.0%	✓
Asphalt Saturation (VFA)	79.5%	65-75%	X (Slightly Exceeds) *

All its Marshall indices fall within the optimal range specified by the standard, with stability exceeding that of traditional empirical mix proportions and orthogonal test results. Compared to traditional methods, this model reduces the number of test trials by over 60%, offering distinct advantages in optimization efficiency and performance. Sensitivity analysis based on Gaussian processes was conducted, and the results of first-order and total effect indices calculated using the Monte Carlo method indicate that the asphalt-to-aggregate ratio has the most significant impact on voids and saturation; the passing rate of 4.75mm plays a dominant role in stability and VMA; and the passing rate of 0.075mm primarily affects the aggregate voids. Bayesian optimization successfully found an excellent mix proportion close to standard requirements, with comprehensive performance superior to that of traditional empirical methods and orthogonal test results. The testing efficiency is high, converging within the complex design space in just 45 trials (including initial trials), avoiding blind trial and error. The asphalt content is reduced by 0.15% compared to the orthogonal test scheme, saving approximately 14 yuan per cubic meter of mixture while offering superior performance. Sensitivity analysis revealed key influencers, with asphalt content and filler content identified as core control parameters, providing a focus for construction quality control. The final recommended BO optimal mix proportion can be directly used for production mix verification, or further fine-tuned based on sensitivity results to meet all standard requirements.

### 5.3. Discussion

In each iteration, BO calculates the acquisition function (such as Expected Improvement (EI)) based on the predicted mean and uncertainty of unknown points from the current model, automatically balancing "exploitation" (sampling in regions with known high potential) and "exploration" (sampling in regions with high uncertainty). This maximizes the information gained from each trial, avoiding blind trial and error. In the AC-13C case, an optimal mix proportion with an 8.1% improvement in comprehensive score was obtained in just 20 iterations (plus 25 initial data points). In contrast, traditional orthogonal tests (e.g., 4 factors at 3 levels) require at least 9 trials but can only examine discrete levels, making it difficult to find the global optimum; full factorial tests (4 factors at 5 levels each) would require 625 trials, which is almost infeasible. The number of trials required by BO grows relatively slowly with problem dimensionality, offering a pronounced advantage in high-dimensional problems. After each new trial result is added to the dataset, the model is immediately updated, gradually approaching the true function. This "learning" closed-loop process allows BO to converge quickly without requiring a predefined complete test plan.

The characteristics of the Gaussian process model determine its robustness and generalization ability: GP not only predicts the mean but also the prediction variance, which is crucial for handling experimental errors (such as random errors in Marshall tests). During the iteration process, points with high uncertainty are prioritized for exploration, thereby reducing the risk of model misjudgment. For example, when the predicted variance in a certain region is too large, BO actively samples to reduce uncertainty, avoiding premature entrapment in a local optimum. By introducing a noise term (WhiteKernel), GP can automatically estimate the level of experimental noise, allowing the model to capture underlying

trends if the data has some fluctuation. In the AC-13C case, cross-validation  $R^2$  values were all above 0.88, indicating that the model has good performance for random errors in the training data. Through active learning strategies (such as selecting high-uncertainty points), BO ensures uniform exploration across the entire design space, thereby improving the model's generalization ability across the entire space.

## 6. Conclusion and Outlook

### 6.1. Main Conclusions

The Bayesian optimization-Marshall test joint optimization model constructed in this study achieved data-driven, small-sample, and multi-global optimization of asphalt mixture proportioning. The model accurately fitted the nonlinear relationship between mixture proportioning and performance using Gaussian processes, efficiently guided the test direction via the Expected Improvement acquisition function, and ultimately obtained a Pareto-optimal solution set that met specification requirements. The study demonstrates that this method significantly reduces testing costs and shortens the design cycle while ensuring road quality, possessing clear theoretical significance and engineering practical value.

### 6.2. Innovations

This study deeply integrates Bayesian optimization with the classic Marshall test to establish an active learning closed-loop optimization. It breaks through the limitations of traditional proportioning design, which relies on experience and extensive testing. By proposing an optimization form targeting multi-indicator deviations and outputting a Pareto-optimal solution set, it provides flexible, scientific, and reliable proportioning schemes for engineering decision-making. Under the premise of ensuring optimization accuracy, the number of physical tests is significantly reduced, making it suitable for rapid proportioning design at the construction site.

### 6.3. Outlook and Suggestions

Currently, only volumetric and basic mechanical indices are optimized; future work can incorporate performance indices such as Semi-Circular Bending (SCB) cracking resistance and Hamburg rutting to construct high-dimensional GP models. For ultra-high-dimensional design variables, sensitive variables can first be screened using a genetic algorithm, followed by fine optimization using Bayesian optimization.

It is suggested to develop a lightweight "Asphalt Intelligent Design Assistant" based on Python/Streamlit, using specification libraries and GP models, to achieve a one-click service from inputting targets to recommending formulas.

## References

- [1] Liang Yongning. Investigation on Mix Proportion and Marshall Test of SMA-13 Asphalt Mixture for Freeway Pavement [J]. China Highway, 2025, (16): 108-110.
- [2] Shu Wei. Discussion on Marshall Test of Asphalt Mixture for Thick Pavement [J]. Subgrade Engineering, 1998, (02): 45.
- [3] Yingcheng L, Weiguang Z, Yongli Z, et al. Mechanical Property Evaluation for Steel Slag in Asphalt Mixture with Different Skeleton Structures Using Modified Marshall Mix Design Methodology [J]. Journal of Materials in Civil Engineering, 2022, 34(1): DOI:10.1611/(ASCE)MT.1943-5533.0004014.

- [4] Xu Jianbin. Comprehensive Report on Marshall Stability Test of Asphalt Mixture [J]. Sichuan Cement, 2020, (08): 341-342.
- [5] Hao Shuwei. Influence of Aggregate Gradation on Marshall Test Results of Asphalt Mixture [J]. Transport World (Construction & Maintenance. Machinery), 2010,(01): 200-202.
- [6] Song M, Chen Z, Xia J, et al. Prediction of post-dryout heat transfer based on physics-embedded machine learning with Bayesian optimization algorithm [J]. Progress in Nuclear Energy, 2026, 194: 106263-16263. DOI:10.1016/J.PNUCENE.2026.106263.
- [7] "Technical Specifications for Construction of Highway Asphalt Pavement" (JTG F40-2004)
- [8] Yu Liwu. Civil Engineering Materials [M]. Nanjing Southeast University press: 202305.
- [9] Zhu Mingxuan, Li Wenkai, Shao Jinggan, et al. Study on Road Performance of Asphalt Mixtures with Different Gradation under Different Loading Conditions [J]. Henan Science, 2026, 44(01): 98-107.
- [10] Nie Guihai. Theory, Design Method and Performance Evaluation of Flowing Asphalt Mixture [D]. Guangzhou University, 2025. DOI:10.27040/d.cnki.ggzdu.2025.000033.
- [11] Xu Hanyong, Xu Wen, Lanhao. Study on Correlation between Stability and Dynamic Stability of Asphalt Mixture [J]. Transportation Science and Technology, 2025,06): 16-20.
- [12] Feng Jun. Research on the Influence of Gradation Segregation on Asphalt Mixture Performance and Its Quality Control [D]. Chang'an University, 2024. DOI:10.26976/d.cnki.gchau.2024.001480.
- [13] Zhu Zehua. Research on Noise Reduction Function and Long-term Acoustic Performance of Thermoplastic Polyurethane/Crumb Rubber Composite Modified Asphalt Pavement [D]. Yangzhou University, 2025. DOI:10.27441/d.cnki.gyzdu.2025000004.
- [14] Qi Tingliang. Study on Fatigue Properties of Warm Mix Asphalt Mixture with High RAP Content [D]. Chongqing Jiaotong University, 2025. DOI:10.27671/d.cnki.gcjtc.2025.001718.5]
- [15] Chen Wenyi. Research on Aggregate Gradation Optimization of AC-13 Asphalt Mixture Based on Data Envelopment Analysis [D]. Jilin Jianzhu University, 2024. DOI:10.27714/d.cnki.gjljs.2024.000172.
- [15] Department of China Journal of Highway and Transport. Review of Academic Research on Chinese Subgrade Engineering - 2025 [J]. China Journal of Highway and Transport, 2025, 38(12): 1-153. DOI:10.19721/j.nki.1001-7372.2025.12.001.
- [16] Si, X. W. Study on Damage Characteristics of Grouting Semi-rigid Pavement Materials under High Temperature and Heavy Load Conditions [henyang Jianzhu University, 2024. DOI: 10.27809/d.cnki.gsjc.2024.000839.
- [17] Jiang, B. Study on Road Performance of Composite Fiber Asphalt Mixture in High Humidity and Saline Environment [D]. Cngqing Jiaotong University, 2024. DOI: 10.27671/d.cnki.gcjtc.2024.000185.
- [18] Liao, Y. Q. Mix Design and Performance Research of Surface Modified Steel Slag Asphalt Mixture in Hot and Humid Areas [D]. Chongqing Jiaotong University, 2025. DOI: 10.27671/d.cnki.gcjtc.2025.001748.
- [19] Niu, H. Application of Bayesian Optimization in Parameter Combination Screening of Functional Textiles [J]. Dyeing Technology, 2025, 47(09): 29-31.
- [20] Chu, Y. L.Chen, H. K., Wang, J. Q., et al. Harmonic Source Modeling of Complex Signals Based on Multivariate Gaussian Process Regression and Spatial Clustering [J]. Transactions of China Technical Society, 2025, 40(23): 7621-7634. DOI: 10.19595/j.cnki.1000-6753.tces.242151.
- [21] Lin, S. Y., Zou, D. Performance Analysis of Two-tier Heterogeneous Cellular Networks Based on Generalized Matern Hard-core Point Process [J]. Application of Electronic Technique, 2025, 51(06): 54-58. DOI: 10.16157/j.issn.028-7998.246025.