Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine-learning prediction of the Superpave high-temperature grade of SBS-modified binders from penetration and softening point, with nomographs for hot climates

mustafa kadhim

mustafa.amoori@s.uokerbala.edu.iq

University of Kerbala

Shakir Al-Busaltan

University of Kerbala

Hayder Hawesah

University of Huddersfield

Research Article

Keywords: SBS polymer, High-temperature performance grade, Penetration, Softening point, Machine learning, Gaussian process regression, Nomograph

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11669573/v1

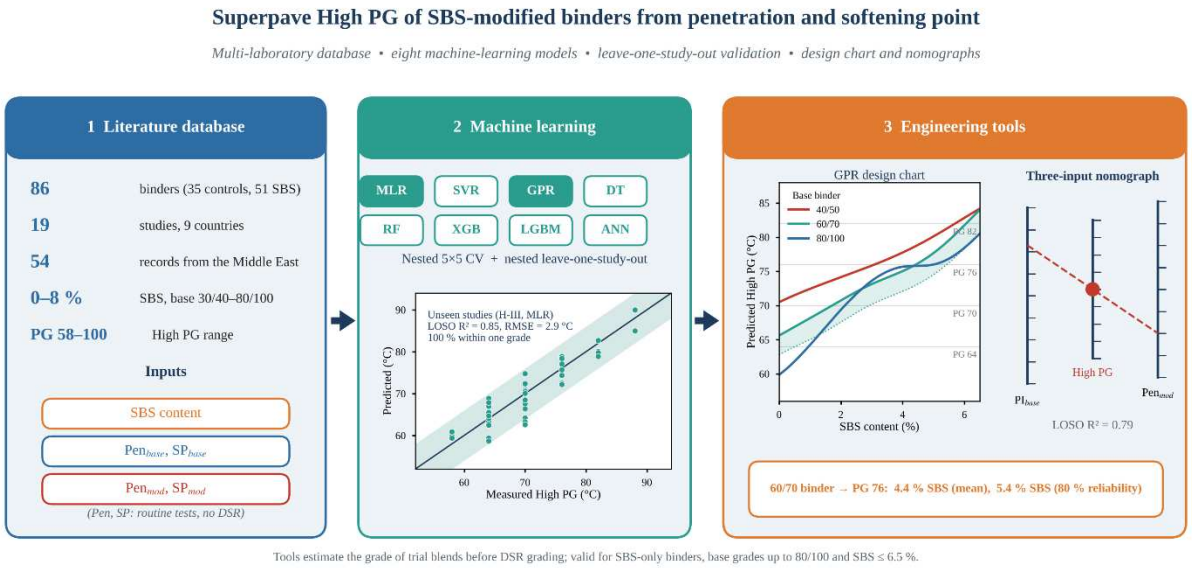
License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Machine-learning prediction of the Superpave high-temperature grade of SBS-modified binders from penetration and softening point, with nomographs for hot climates

Highlights

- A filtered database of 86 SBS and control binders with Pen ≤ 100 dmm was built
- Eight ML models predicted High PG from SBS content, Pen and SP (CV R^2 up to 0.85)
- Pen and SP, before and after modification, gave a LOSO R^2 of 0.85 for High PG
- A three-input PI–Pen–SBS nomograph predicted unseen studies with LOSO R^2 = 0.79
- Nomographs and heat maps estimate the High PG of trial SBS blends before DSR grading



Abstract

Styrene–butadiene–styrene (SBS) is the polymer most widely used to increase the high-temperature performance grade (High PG) of bitumen in hot regions such as the Middle East, yet many laboratories there still specify binders by penetration (Pen) and softening point (SP). A literature database of 86 SBS-modified and control binders from 19 studies was compiled for base binders of penetration grades up to 80/100 (Pen ≤ 100 dmm), covering PG 58 to PG 100 and up to 8 % SBS. Eight machine-learning algorithms were compared across three scenarios, using SBS content with the Pen, SP, and penetration index of the base binder (H-I), the modified binder (H-II), or both (H-III), and were assessed using hold-out testing, nested 5 × 5-fold cross-validation (CV), nested leave-one-study-out (LOSO) validation, and practitioner

baselines (no change, and the base grade plus a fitted SBS slope). The best models achieved CV R^2 = 0.85, 0.70, and 0.84 (CV RMSE = 3.8, 4.1, and 3.0 °C), whereas assuming the base-binder grade yielded negative R^2. For the SBS-modified binders alone, the best H-III model gave a LOSO RMSE of 2.7 °C, against 5.7 °C for the base grade plus a fitted SBS slope. Consistency before and after modification transferred best across studies (LOSO R^2 = 0.85, RMSE = 2.9 °C), and SHAP identified SBS content and modified-binder penetration as key inputs. The effect of SBS was 3.3–3.9 °C per 1 % across pooled, within-study, and mixed-model estimates, but about 2.0–2.5 °C per 1 % for a 60/70 binder in the design range. A Gaussian-process design chart indicates mean dosages of about 1.7, 4.4, and 6.1 % SBS to raise a 60/70 binder to PG 70, 76, and 82, or 2.9 and 5.4 % for PG 70 and PG 76 at 80 % reliability; contents above 6.5 % SBS are not supported by the data. Three-input nomographs and two-dimensional heat maps based on SBS content with PI_{base} and Pen_{mod} (LOSO R^2 = 0.79) or with SP_{base} and SP_{mod} (LOSO R^2 = 0.75) estimate the High PG of trial blends before DSR grading.

Keywords: SBS polymer; High-temperature performance grade; Penetration; Softening point; Machine learning; Gaussian process regression; Nomograph

1. Introduction

The most common distress in asphalt pavements in hot climates is permanent deformation (rutting), although rutting resistance is governed mainly by the aggregate skeleton and the volumetrics of the mixture, the binder contributes substantially through its stiffness and elasticity at high service temperatures [1,2]. The Eastern Mediterranean and the Middle East are warming nearly twice as quickly as the world, at approximately 0.45 °C per decade, and summer heatwaves are expected to be longer and more intense [3]; in Iraq, rutting has been observed on highways within a year of construction [1]. Climate-based binder specifications developed for the region require PG 76-10 in most of the United Arab Emirates and PG 70-10 in its cooler areas [4], and PG 70-10 or PG 76-10 across more than 70 % of Iraq [5]. Local refinery binders, though, are made as 40/50 and 60/70 penetration grades that normally grade only as PG 64 or PG 70 [5–7]; the Iraqi specification continues to define binders by penetration and viscosity [8]. Since the design low pavement temperatures in Iraq and the Gulf are mild (PG xx-10) [4,5], the high-temperature performance grade (High PG) is the criterion that most commonly determines which binder to use, and polymer modification is often necessary to achieve it [5–7]. The current research paper thus focuses on the High PG.

The most common binder modifiers are styrene-butadiene-styrene (SBS) block copolymers [9,10]. When mixed, the polybutadiene blocks absorb light maltene fractions and swell. When

the concentration exceeds a critical level, the polymer-rich fraction forms a continuous network, physically cross-linked by polystyrene domains, increasing stiffness and elasticity at high temperatures [9–11]. The magnitude of the effect depends on polymer content, compatibility with the base bitumen, and polymer architecture, i.e., linear or radial (branched) SBS, styrene and vinyl contents, and molecular weight [12–15]. Phase inversion to a continuous polymer-rich phase is usually reported at about 4–7 % SBS, depending on the compatibility of the base bitumen [16,17], whereas storage stability and the viscosity at mixing and compaction temperatures limit the practical content to about 6–7 % for many binders [10,18].

In the Superpave system, the High PG is the highest temperature, in 6 °C increments, at which the dynamic shear rheometer (DSR) parameter $|G^*|/\sin\delta$ of the original binder is at least 1.00 kPa and that of the rolling thin-film oven (RTFO) residue is at least 2.20 kPa [19–21]. For polymer-modified binders, $|G^*|/\sin\delta$ captures only part of the elastic contribution of the polymer network, which motivated the multiple stress creep recovery test [22,23]; nevertheless, the M 320 grade remains the basis of binder selection in most specifications in the region. The multiple stress creep recovery (MSCR) grade of AASHTO M 332 [24], based on AASHTO T 350 [25], is better suited to polymer-modified binders; the M 320 $|G^*|/\sin\delta$ grade was nevertheless chosen as the target because the climate-based binder specifications of the region are written in M 320 grades [4,5] and because MSCR results are reported far less often than M 320 grades in the studies available for the database (see also Section 8.3). In the region, where the penetration (Pen) and softening point (SP) tests [26,27] are the standard tools of quality control, DSR equipment is still unavailable in many laboratories. These empirical tests cannot be directly associated with the viscoelastic properties of bitumen [28]; binders of the same penetration grade in other sources may vary significantly in their rheology [29]; and the correlations that have been made between Pen, SP, or the penetration index (PI) and Superpave parameters have been obtained largely on unmodified binders or on one or a few base binders in a single laboratory [28,30]. Nor do they indicate the SBS content needed for a binder of a given penetration grade to reach a target grade.

Machine learning (ML) is increasingly used in pavement engineering [31,32] to approximate responses that are difficult to measure, such as the dynamic modulus [33], rut depth [34], and fracture resistance [35,36] of asphalt mixtures; the viscoelastic and rheological properties of modified binders [37,38] and their high-temperature failure temperature [39]; and the SBS content, Pen, and SP of SBS-modified asphalt [40]. For SBS-modified binders, however, to the

authors' knowledge (based on the search described in Section 2.1), no paper has (i) predicted the High PG based on the polymer content and routine consistency tests, (ii) compared a large group of algorithms on a multi-laboratory database covering a multiplicity of base binders and SBS products, (iii) tested the transferability of the models to unseen laboratories, or (iv) presented the results in charts, nomographs, and heat maps that do not require software. Riyad et al. [39] predicted the high- and intermediate-temperature performance of modified binders from their own experiments, and Zhong et al. [40] modeled the Pen and SP of SBS-modified asphalt from its composition; neither used multi-laboratory data, validated the models on unseen studies, or translated them into design charts and nomographs.

This study addresses these gaps. It (i) compiled and sieved a literature database of SBS-modified binders that report the Superpave High PG, along with the SBS content and the Pen and SP of the base and/or modified binder; (ii) analyzed the statistical relationships between these variables and the High PG; (iii) developed and tuned eight ML algorithms with three input scenarios (SBS content and binder consistency prior to modification, after modification, and prior to and after change) and tested them using hold-out testing, nested repeated cross-validation, nested leave-one-study-out validation, and practitioner baselines; (iv) quantified the importance of the inputs using in-sample, out-of-fold, and grouped permutation importance and SHAP; (v) provided an explicit Gaussian process regression (GPR) formula and a design chart giving the mean and the reliability-based SBS content needed to reach PG 70, 76, and 82; and (vi) developed three-input alignment nomographs and two-dimensional heat maps of the SBS content and binder consistency before and after modification, and examined whether the polymer structure should be an additional input.

The scope of the study is defined by the sieved database (Section 2.1). It includes 86 binders (35 unmodified controls and 51 SBS-modified binders) published in 19 studies across 9 countries, with 54 from the Middle East (Iraq 27, Turkey 15, Iran 6, United Arab Emirates 4, and Saudi Arabia 2; Turkey is counted as Middle Eastern) and 32 (37 %) from India, China, Brazil, and Malaysia, which are not specific to hot climates. The base binders are penetration grades 30/40 to 80/100 (Pen_{base} = 33–97 dmm, SP_{base} = 45.3–55.6 °C). Softer grades, like 160/220, are not used in pavements in high-temperature regions and were excluded. The SBS content ranges from 0.5 % to 8 %. The High PG of the unmodified binders is 58 to 76, and the High PG of the modified binders is 64 to 100. The models, charts, and nomographs can be used within these ranges and only for binders modified with SBS alone (no co-modifiers or other

polymers); the design chart and the nomographs are further restricted to SBS contents of up to 6.5–7 %, above which the data are sparse (Sections 6.2 and 7.2).

2. Materials and methods

The research program is summarized in Fig. 1. It comprised compiling and sieving a literature database, formulating three input scenarios, developing, tuning, and validating eight ML algorithms, statistically comparing and interpreting the results, and deriving a prediction formula, a design chart, and nomographs. All computations were carried out in Python 3.13 using scikit-learn 1.9 [41], XGBoost 3.4, LightGBM 4.7, statsmodels 0.15, SHAP 0.54, NumPy 2.5, and pandas 3.0; all random splits and estimators used the fixed seed 42 (outer CV repetitions r = 0–4 used seeds 42 + r), so that the CV and LOSO splits can be reproduced. The database, code, and fitted models are provided as supplementary material.

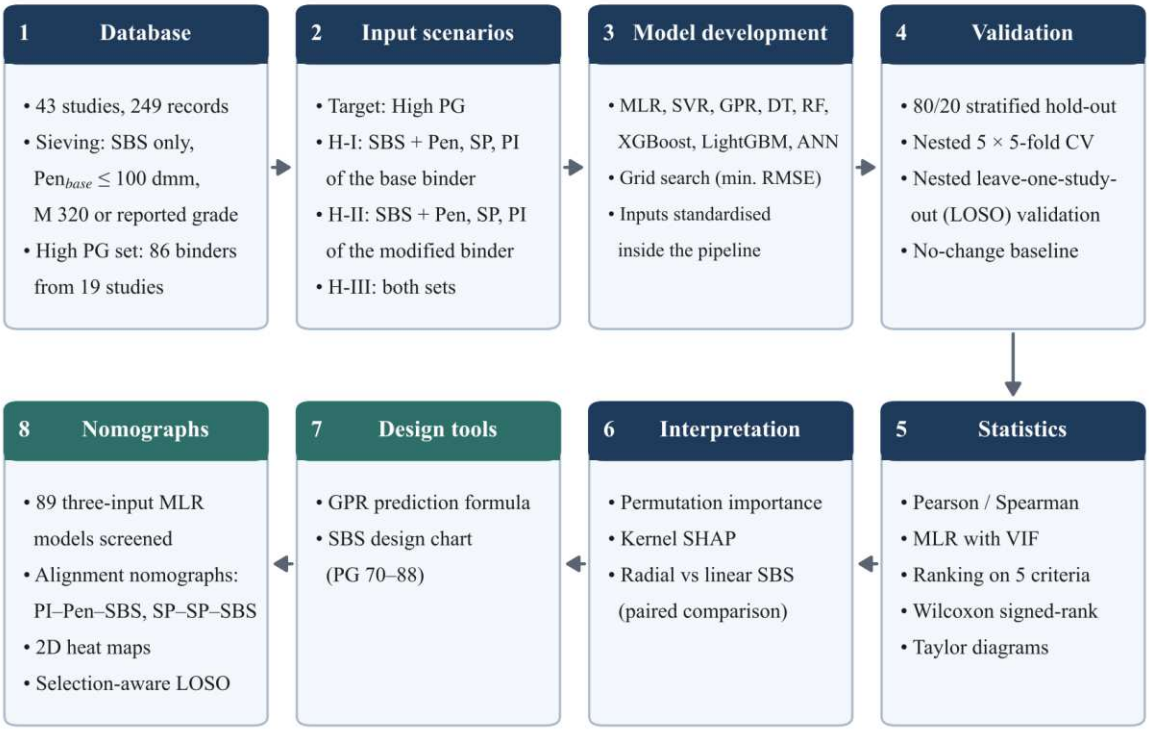


Fig. 1. Framework of the research program for predicting the High PG of SBS-modified binders: data (step 1), modeling and validation (steps 2–4), analysis (steps 5–6) and practical tools (steps 7–8).

2.1. Database compilation and data sieving

Experimental studies on SBS-modified binders were identified from the authors' literature collection and from searches of Scopus, Web of Science, ScienceDirect, and Google Scholar, supplemented by regional university journals (Iraqi Academic Scientific Journals portal) and agency reports [databases, search dates and numbers of hits to be confirmed by the authors]. The search combined the terms ("SBS" OR "styrene–butadiene–styrene") AND ("performance

grade" OR "Superpave" OR "dynamic shear rheometer" OR "G*/sin δ") AND ("penetration" OR "softening point") in titles, abstracts, and keywords, covered journal articles, conference papers, theses, and reports in English published from 2000 to 2025, and was not restricted to open-access sources. Studies were included when they reported, for at least one binder, the SBS content, the Pen or SP of the base or modified binder, and a High PG or a high-temperature DSR result; studies on mixtures only, on SBS used only together with other modifiers, or without numerical binder data were excluded. The identification, screening, and inclusion steps follow a PRISMA-type flow [42], whose sieving stage is given in Table 1. Every candidate study was read in full, and 43 studies reported binder data that could be extracted. For each binder, the base binder and its penetration grade, the SBS content (% by weight of base binder), the SBS architecture where reported, the Pen [26] and SP [27] of the base and the modified binder, and the graded or continuous High PG of the base and the modified binder were recorded, along with the table or figure from which each value was taken. Values read from labeled bar charts or digitized from scatter plots were flagged in the database; in the High PG data set this concerns 5 records from two studies [43,44], and removing them changed the LOSO R^2 of the best models by at most 0.03 (Section 4.5). The base-binder data of one study [45] were checked against a companion paper by the same authors on the same base binder [46], which led to a correction of its base-binder penetration from 50.9 to 58.0 dmm; the results in Sections 3–7 were computed with the original value, and the effect of the correction (at most 0.012 in R^2) is shown in Section 4.5. In total, 249 binder records were extracted. The 19 studies that provide the High PG data set are cited here [6,7,13,14,30,43–45,47–57]; their numbers of records, countries, base-binder grades, SBS types, and grading bases are listed in Supplementary Table S1 (sheet "Sources" of the supplementary database).

The records were then sieved using the rules in Table 1. Only binders modified with SBS alone were retained: binders containing oils, acids, stabilizers, compatibilizers, recycled binder, other polymers, diblock SB copolymers, and commercial binders of undisclosed SBS content were removed, so that SBS content and binder consistency are the only modeling inputs. Linear and radial SBS were treated as one modifier, and their difference was examined separately in Section 7.5. Finally, base binders softer than penetration grade 80/100 (Pen_{base} > 100 dmm) were removed because such binders are not used for pavements in hot climates and their very low grades would dominate the variation in High PG. A High PG was accepted when it was determined from the original and the RTFO-aged binder following AASHTO M 320 [19] or reported as a grade by the authors (57 and 29 records, respectively). The reported grades could

not all be checked for the RTFO criterion or for the edition of M 320 used; their effect was therefore tested by repeating the validation on the 57 records with full M 320 grading (Section 4.5); upper failure temperatures measured on the original binder only were kept in the database because their Pen and SP values are useful for the consistency relationships, but were not used as High PG. Unmodified control binders were kept with an SBS content of 0 %, for which the consistency after modification equals that of the base binder. Most source studies report a single set of Pen and SP for the unmodified binder, and whether the controls were subjected to the same blending history (temperature, shear, and time) as the SBS blends is rarely stated, so $Pen_{mod} = Pen_{base}$ for the controls is an approximation. The controls make up 35 of the 86 records; the number of SBS-modified records in each scenario is given in Section 2.3, and the accuracy for the SBS-modified binders alone is reported in Section 4.4. Of the 72 records removed in step 5 of Table 1, 22 had only upper failure temperatures of the original binder and 50 had no High PG at all.

Table 1. Sequence of data-sieving rules and number of binder records and studies remaining.

Step	Sieving rule	Records (studies)
1	Binder records extracted from studies reporting SBS-binder data	249 (43)
2	Remove binders with co-modifiers, other polymers, SB diblock copolymers, commercial PMBs of undisclosed composition, or neither a High PG nor consistency data after modification	169 (32)
3	Remove binders whose SBS content is not reported for the individual binder (the study gives a range or a nominal value only)	168 (32)
4	Remove base binders softer than 80/100 (Pen_{base} > 100 dmm)	158 (32)
5	High PG data set: full AASHTO M 320 grading (original + RTFO) or reported grade	86 (19)

The scope and ranges of the resulting High PG data set, which define the applicability limits of the models, charts, and nomographs, are summarized in Table 2.

Table 2. Scope and ranges of the High PG database (applicability limits of the models, charts and nomographs).

Property	Value
Studies (records)	19 (86)
Publication years	2010–2025
Countries (records)	Iraq (27), India (18), Turkey (15), China (11), Iran (6), UAE (4), Saudi Arabia (2), Brazil (2), Malaysia (1)
Unmodified controls / SBS-modified binders	35 / 51
SBS content (% by weight of base binder)	0.5–8 (median 4.5)
SBS architecture (modified binders)	linear (25), not reported (19), radial (7)
Penetration grades of the base binders	30/40 to 80/100
Pen_{base} (dmm) / SP_{base} (°C) / PI_{base}	32.8–97.0 / 45.3–55.6 / −2.19–0.29
Pen_{mod} (dmm) / SP_{mod} (°C) / PI_{mod} (SBS binders)	33.5–93.0 / 47.0–104.4 / −0.40–4.73
High PG of unmodified binders	PG 58 to PG 76
High PG of SBS-modified binders	PG 64 to PG 100
Grading basis (records)	AASHTO M 320 original + RTFO (57); grade reported by the authors (29)

Consistency ranges refer to the records of the High PG data set for which the property is known; PI computed by the Pfeiffer–Van Doormaal relationship.

2.2. Input variables and grade conversion

The PI of the base and the modified binder was calculated from Pen and SP by the Pfeiffer–Van Doormaal relationship [58]:

$$PI = (20 - 500A) / (1 + 50A), \quad A = [\log(800) - \log(Pen)] / (SP - 25) \tag{1}$$

where Pen is the penetration at 25 °C (dmm) and SP the softening point (°C). The relationship assumes a penetration of about 800 dmm at the softening point, which holds approximately for unmodified bitumen but not for SBS binders, whose SP is raised by the elastic polymer network [16,59]; PI_{mod} is therefore not a valid temperature-susceptibility index for SBS binders. Because PI is also a deterministic function of Pen and SP (the cause of the large VIFs discussed in Section 3.3), it was used only as an engineered input feature, and its importance is interpreted jointly with Pen and SP (Section 5); in the nomograph of Section 7.2, it refers to the unmodified base binder only. Where only continuous grades were reported, the High PG was taken as the highest standard grade not exceeding the continuous high temperature $T_{c,H}$ [60]:

$$\text{High PG} = 6 \cdot \lfloor (T_{c,H} - 4)/6 \rfloor + 4 \tag{2}$$

A continuous $T_{c,H}$ was available for 40 of the 86 records, so the grade, rather than $T_{c,H}$, was used as the common target. Because the grade is discrete in 6 °C steps, rounding alone introduces an error with a standard deviation of about $6/\sqrt{12} \approx 1.7$ °C, which is a floor for the RMSE of any model of the grade; the RMSE values of Section 4 should be read against this floor.

The grade of the unmodified binder of the same study and base binder (High PG_{base}) was attached to every record for the no-change baseline (Section 2.5); it was not used as a model input.

2.3. Prediction scenarios

Three scenarios were formulated (Table 3). Scenario H-I represents the design stage in a laboratory that specifies binders by penetration grade: only the SBS content and the Pen, SP, and PI of the base binder are known. Scenario H-II represents the verification stage, when a trial blend has been prepared and its Pen and SP measured. Scenario H-III combines both sets of information. Each scenario used all sieved records for which its inputs and the High PG were available, so the number of records differs across scenarios. To separate the effect of the inputs from that of the records, the comparison was repeated on the 55 records common to all three scenarios (Section 4.4).

Table 3. Input variables of the three High PG scenarios.

Scenario	Input variables	n (studies)	SBS-modified / controls
H-I	SBS, Pen_{base}, SP_{base}, PI_{base}	85 (18)	50 / 35
H-II	SBS, Pen_{mod}, SP_{mod}, PI_{mod}	56 (19)	21 / 35
H-III	SBS, Pen_{base}, SP_{base}, PI_{base}, Pen_{mod}, SP_{mod}, PI_{mod}	55 (18)	20 / 35

SBS = SBS content (% by weight of base binder); subscripts base and mod denote the binder before and after modification (for unmodified controls both are equal).

2.4. Machine-learning algorithms

Eight algorithms from four families were compared, as is common practice for small experimental data sets of asphalt materials [37,61]: multiple linear regression (MLR) as a statistical benchmark; support vector regression (SVR) [61–63] and GPR [38,64,65] as kernel methods; a single decision tree (DT) [37] and the ensembles random forest (RF) [66,67], XGBoost [67–69], and LightGBM [70,71] as tree-based methods; and a feed-forward artificial neural network (ANN) [33,40]. MLR estimates $\hat{y} = \beta_0 + \Sigma\beta_j x_j$ by ordinary least squares. SVR with a radial basis function kernel fits $f(x) = \Sigma_i(\alpha_i - \alpha_i^*) \exp(-\gamma\|x_i - x\|^2) + b$ while ignoring deviations smaller than ε. GPR places a Gaussian-process prior on the response and returns the posterior mean $\mu_* = k_*^T(K + \sigma_n^2 I)^{-1}y$ and variance $\sigma_*^2 = k(x_*, x_*) - k_*^T(K + \sigma_n^2 I)^{-1}k_*$; a scaled Matérn kernel with a white-noise term was used, with the smoothness $\nu \in \{0.5, 1.5, 2.5\}$ selected by grid search and the other kernel parameters fitted by maximizing the log-marginal likelihood. RF averages 300 trees grown on bootstrap samples, whereas XGBoost and LightGBM build 300 trees sequentially using the gradient of a regularized loss [69,71]. The ANN was a multilayer perceptron with one or two hidden layers, trained using the limited-memory BFGS algorithm, which is appropriate for small data sets. The tree-based models cannot extrapolate beyond the range of the training targets. The kernel models and the ANN give smooth response surfaces, but smoothness does not make extrapolation reliable: away from the data, a Matérn GPR reverts towards the prior (training) mean, the RBF-SVR flattens, and the ANN can extrapolate arbitrarily [64,72]. At the 6–8 % SBS edge of the database, which matters for the design chart (Section 6.2), the GPR is therefore expected to underestimate rather than overestimate the grade gain, and this region is flagged in the chart. Table 4 shows the tuned hyperparameters.

Table 4. Optimal hyperparameters obtained by grid search (2 × 5-fold CV on the training set); they were used for the hold-out test and the final models, whereas the CV and LOSO estimates re-tuned them inside every outer fold (Section 2.5).

Model	H-I	H-II	H-III
SVR	C = 100, γ = 0.5, ε = 1.0	C = 100, γ = 0.01, ε = 1.0	C = 1000, γ = 0.01, ε = 1.0
GPR	ν = 0.5	ν = 1.5	ν = 2.5
DT	depth = 4, leaf = 1	depth = 4, leaf = 1	depth = 3, leaf = 5

Model	H-I	H-II	H-III
RF	depth = None, feat. = 1.0, leaf = 1	depth = None, feat. = 1 (one feature), leaf = 1	depth = None, feat. = 0.5, leaf = 1
XGBoost	η = 0.1, depth = 4, λ = 10, sub. = 0.8	η = 0.03, depth = 2, λ = 1, sub. = 0.8	η = 0.03, depth = 2, λ = 10, sub. = 0.8
LightGBM	η = 0.1, leaves = 4, λ = 1, min. child = 3	η = 0.03, leaves = 8, λ = 10, min. child = 5	η = 0.03, leaves = 4, λ = 10, min. child = 5
ANN	(8), relu, α = 0.001	(16), relu, α = 1.0	(8), tanh, α = 1.0

MLR has no hyperparameters. η = learning rate; λ = L2 regularization; feat. = fraction of features per split (1.0 = all features) or, for an integer, the number of features per split; sub. = row subsample; ANN: hidden-layer sizes, activation, L2 penalty α. None = unlimited depth.

2.5. Validation protocol

For each scenario, records were split into a training set (80 %) and a hold-out test set (20 %), stratified by grade (grades with only one record, such as PG 94 or PG 100 where applicable, were merged with the nearest grade class for stratification). For the hold-out test and for the final models used in the importance analysis and the prediction formula, hyperparameters were tuned on the training set only using an exhaustive grid search with 2 × 5-fold repeated cross-validation [73,74], minimizing the root mean square error (RMSE). For MLR, SVR, GPR, and the ANN, inputs were standardized within a pipeline so that scaling was learned from the training folds only. Because the hold-out sets are small (11–17 records) and a single split yields a noisy estimate [75,76], two additional procedures were applied. Both were fully nested, so that records used to estimate accuracy never influenced the choice of hyperparameters [76,77]: in every outer fold, the entire grid was searched again on the outer training records only; the combination with the lowest mean RMSE over the inner folds was selected; the model was refitted with it on all outer training records; and only then were the outer test records predicted (Fig. 2). First, nested 5 × 5-fold repeated cross-validation (CV) was performed on all records of each scenario, with random inner 5-fold CV for tuning; the mean and standard deviation over the five repetitions are reported, and the repeat-averaged out-of-fold predictions were used for the Taylor diagrams, residual analysis, and statistical tests. Second, in a nested leave-one-study-out (LOSO) validation, every study was withheld in turn and predicted by a model tuned and trained on all other studies, so that records sharing a laboratory, base binder, or SBS product never appear in both the training and the validation set [77,78]. Within each LOSO fold, hyperparameters were tuned by a 5-fold inner CV grouped by study (GroupKFold), so that the inner folds also mimic prediction for an unseen study. The LOSO R^2 and RMSE were computed from the pooled held-out predictions of all studies; because a pooled R^2 is influenced by the largest studies and by the between-study range of grades, the median and interquartile range of the per-study RMSE are also reported (Section 4.5). LOSO measures how well a model transfers to a laboratory, base binder, SBS product, and blending protocol that it has never seen, which is the realistic use of a literature-based model. Finally, a no-change baseline,

which predicts the High PG of the modified binder to be equal to that of its base binder ($PG_{mod} = PG_{base}$), was evaluated on all records for which the base grade was known to assess whether the models add information beyond the base grade. Because the no-change baseline is exact for the unmodified controls and ignores the SBS content, two stronger baselines that a practitioner could use were added: a linear model of the SBS content alone, and $PG_{mod} = PG_{base} + b \cdot SBS$, with the slope b fitted within the training folds (equivalent to modeling the grade gain $\Delta PG = PG_{mod} - PG_{base}$ as proportional to the SBS content); both were evaluated with the same 5 × 5-fold CV and LOSO splits. All metrics were also computed for the SBS-modified binders alone.



Fig. 2. Nested leave-one-study-out validation: (a) outer loop, in which each study is held out once; (b) hyperparameter selection inside every outer fold, using a grid search with a 5-fold inner CV grouped by study on the training studies only and the minimum mean inner RMSE as the selection criterion, followed by refitting and prediction of the held-out study.

2.6. Performance metrics and interpretation

Accuracy was quantified with the coefficient of determination, the RMSE and the mean absolute error (MAE), where y_i and $\hat{y}_i$ are the measured and predicted High PG of record i, $\bar{y}$ is the mean measured value and n the number of records:

$$R^2 = 1 - \Sigma(y_i - \hat{y}_i)^2 / \Sigma(y_i - \bar{y})^2 \quad (3)$$

$$RMSE = \sqrt{[(1/n)\,\Sigma(y_i - \hat{y}_i)^2]}, \quad MAE = (1/n)\,\Sigma|y_i - \hat{y}_i| \quad (4)$$

The following metrics are provided in the supplemental results file: the correlation coefficient, mean absolute percentage error, variance accounted for, Willmott index of agreement, and the a10 index [79]. The accuracy of the High PG is discrete in 6 °C steps, so two engineering measures were introduced: the accuracy of the exact grade (i.e., the proportion of records predicted to the nearest standard grade that matches the measured grade) and the within-one-grade accuracy (±6 °C). The models were ranked on five criteria (CV R^2, CV RMSE, CV MAE, LOSO R^2, and the within-one-grade accuracy of the CV predictions, with ties given the minimum rank); because the three CV accuracy measures are correlated, the ranking was also repeated on the LOSO RMSE alone (Section 4.4), and the absolute out-of-fold errors of the most accurate model for each scenario were compared with those of the others using the Wilcoxon signed-rank test [80], with Holm's correction for the seven comparisons in each scenario [81]; because the out-of-fold errors of records from the same study are not independent, these p-values are indicative only. These out-of-fold predictions were also summarized in Taylor diagrams [82].

The contribution of each input was quantified by permutation importance [66,83], computed for every model and scenario as the mean increase in RMSE over K = 20 random permutations of input j, using the models refitted on all records of the scenario:

$$I_j = (1/K)\,\Sigma_k\,[\mathrm{RMSE}(X_{\mathrm{perm},j,k}) - \mathrm{RMSE}(X)] \quad (5)$$

Permutation importance computed on the training data measures how much a fitted model relies on an input rather than its predictive importance, and permuting one of the functionally linked inputs (PI = f(Pen, SP)) creates impossible Pen–SP–PI combinations [84]. The analysis was therefore repeated out of fold (3 × 5-fold CV, with the importance computed on the held-out folds and 10 permutations per fold), with the inputs permuted jointly in three groups: the SBS content, the base-binder consistency $\{Pen_{base}, SP_{base}, PI_{base}\}$, and the modified-binder consistency $\{Pen_{mod}, SP_{mod}, PI_{mod}\}$ [83].

In addition, SHAP values [85] decompose each prediction into additive contributions, $\hat{y}(x) = \varphi_0 + \Sigma_j\varphi_j(x)$, where φ_0 is the average prediction and φ_j the Shapley value of input j. The model-agnostic Kernel SHAP estimator was applied to the best-ranked non-linear model of each scenario with a k-means background of ten centroids. Kernel SHAP assumes independent inputs, so the attributions among the Pen, SP, and PI of the same binder are not separately identifiable; grouped SHAP values (the sum of the SHAP values of a group) are therefore also reported.

2.7. Development of the nomographs

Additive MLR models of the form High PG = $c + f_1(x_1) + f_2(x_2) + f_3(x_3)$ were converted into three-input nomographs. All combinations of three variables selected from the SBS content, the Pen (linear or logarithmic), SP and PI of the base and modified binders, and the SP gain $\Delta SP = SP_{mod} - SP_{base}$ were screened to select the inputs. Of the $C(10,3) = 120$ triplets of these ten candidates, 17 that combined the linear and logarithmic penetration of the same binder or all three SP variables were excluded, and 14 of the remaining 103 that contain ΔSP together with one softening point are linear reparameterizations of a model with both softening points, leaving 89 distinct candidate models. Each candidate model was fitted by ordinary least squares to all records with its three inputs available and validated by 10 × 5-fold CV and LOSO; because OLS has no hyperparameters, non-nested CV with ten repetitions was used here, whereas the nested 5 × 5-fold CV of Section 2.5 was needed only for the tuned ML models. The combination with the greatest LOSO R^2 was retained, and the combination of SP_{base}, SP_{mod}, and the SBS content was selected because SP is the most commonly reported consistency test. The equations were plotted as a two-stage alignment nomogram, with the turning line ungraded [86]. The x_1 scale is superimposed on abscissa 0 with ordinates $s_1 f_1(x_1)$, and the x_2 scale on abscissa 1 with ordinates $s_2 f_2(x_2)$; a straight line through x_1 and x_2 will cut the turning line T, which is placed on abscissa t_1 with modulus k_1:

$$t_1 = s_1/(s_1 + s_2), \quad k_1 = s_1 s_2/(s_1 + s_2) \tag{6}$$

Then a second line between this and the x_3 scale (modulus s_3) passes through the High PG scale, which is at the fraction t_2 of the distance between T and the x_3 scale and has a modulus k_2:

$$t_2 = k_1/(k_1 + s_3), \quad k_2 = k_1 s_3/(k_1 + s_3) \tag{7}$$

where s_1, s_2, and s_3 are the scale moduli, selected to produce scales of similar length; constant offsets shift the scales only up and down. The constructions were verified numerically: for the extreme input combinations and the worked examples, the value read on the High PG scale reproduced Eqs. (10) and (11) exactly (difference < 10^{-9} °C), so the reading precision is limited only by the graduation of the scales. Because a nomograph does not fix any input, the same equations were also plotted as heat maps of the High PG in two dimensions as functions of the two consistency inputs at SBS contents of 2, 4, and 6 %, with the grade boundaries represented as contours and the measured records of the neighboring SBS contents superimposed. Lastly, the out-of-fold residuals of the H-I GPR were grouped by SBS type (linear, radial, not

reported). Separately, binders with linear and radial SBS prepared from the same base binder at the same SBS content within the same study were identified and compared in pairs.

3. Database characteristics and statistical analysis

3.1. Composition of the database

The data set for the High PG is presented in Fig. 3 and Table 5, along with its descriptive statistics. It includes 86 records from 19 studies, 35 of which are unmodified controls and 51 of which are SBS-modified binders with 0.5–8 % SBS (median of the modified binders is 4.5 %). The High PG is measured to be between 58 and 100 °C, and the base binders have softening points of 45.3 to 55.6 °C and penetrations between 30/40 and 80/100 (Table 2). Of the 86 records, 54 (27 Iraqi) come from Middle Eastern studies, so the database is weighted towards hot-climate binders. This sieving step eliminated 160/220 base binders (Section 2.1), increased the lowest High PG of the data set to PG 58, and limited the base binders to hot-climate pavement grades.

The 49 SBS-modified binders with a known base grade improved by 0, 1, 2, 3, 4, 5, 7 grades in 2, 10, 16, 11, 6, 1, 3 cases, respectively (Fig. 3d); the median gain was 2 grades at a median SBS content of 4.5 %, and the median gain per 1 % SBS was 4.0 °C. The spread at a particular SBS content is high: the gains at 4.5 to 5 percent SBS were between 1 and 5 grades. The three binders that gained seven grades (PG 58 to PG 100 with 7 % SBS) come from one study [13], which reported continuous failure temperatures of 100–105 °C; the values agree with the tables of the source, but because the grading basis could not be confirmed as full M 320 grading and the records are influential at the top of the range, the models were also validated without the four binders that gained five grades or more (Section 4.5). The median gain of 4.0 °C per 1 % SBS is a median of ratios for the modified binders alone and is not directly comparable with the regression slopes of Section 3.3.

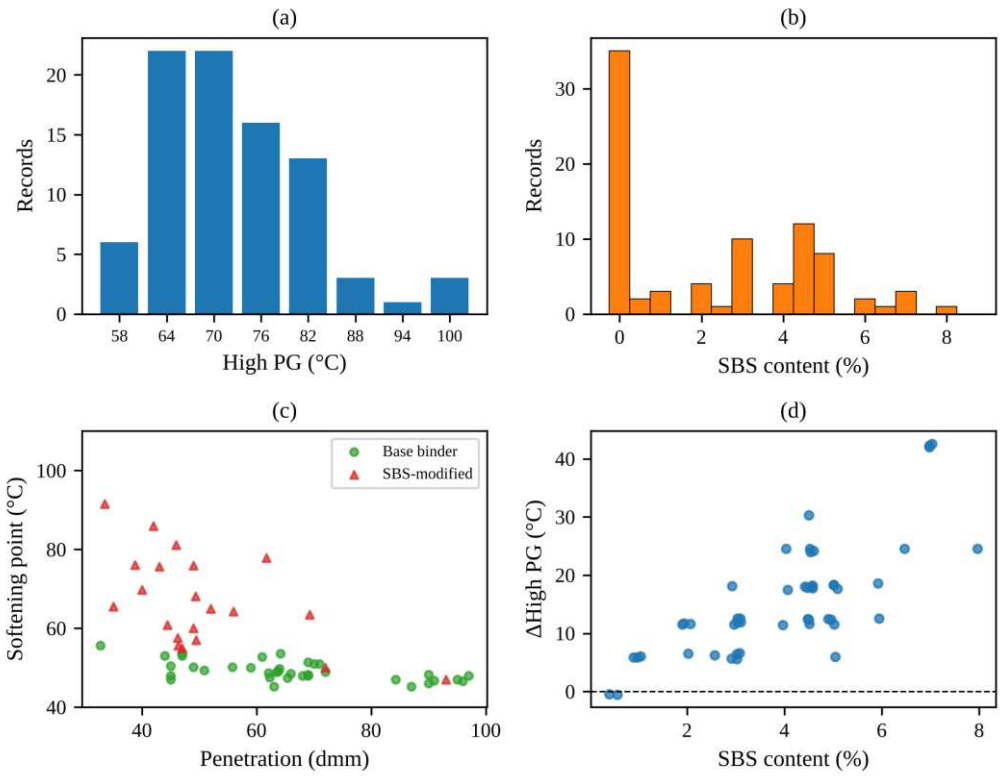


Fig. 3. Composition of the High PG data set: (a) measured High PG; (b) SBS content; (c) Pen and SP of the base and the SBS-modified binders (High PG data set); (d) change of High PG relative to the base binder versus SBS content (points slightly jittered).

Table 5. Descriptive statistics of the variables of the High PG data set.

Variable	n	Min	Max	Mean	SD	Median
SBS (%)	86	0.0	8.0	2.3	2.4	2.0
Pen_{base} (dmm)	85	32.8	97.0	68.8	17.7	64.2
SP_{base} (°C)	85	45.3	55.6	49.1	2.4	48.5
PI_{base}	85	−2.2	0.3	−0.7	0.6	−0.7
Pen_{mod} (dmm)	56	32.8	97.0	60.2	17.4	60.0
SP_{mod} (°C)	71	45.3	104.4	62.8	17.5	53.6
PI_{mod}	56	−2.2	4.7	0.3	1.8	−0.4
High PG_{base} (°C)	84	58.0	76.0	63.5	4.4	64.0
High PG (°C)	86	58.0	100.0	72.5	9.6	70.0

Statistics of the variables after modification include the unmodified controls (Pen_{mod} = Pen_{base}).

3.2. Correlation analysis

The Pearson and Spearman coefficients are listed in Table 6; the Pearson matrix is shown in Fig. 4; and the main relationships are shown in Fig. 5. The strongest correlates of the High PG are the consistency of the modified binder, SP_{mod} ($r = 0.831$) and Pen_{mod} ($r = -0.755$), and the SBS content ($r = 0.799$). All three are significant at $p < 0.001$, and PI_{mod} is also significant ($r = 0.615$). Within the penetration grades covered (30/40–80/100), the marginal correlations between the consistency of the base binder and the High PG of the modified binder are weak (SP_{base}: $r = 0.156$, $p = 0.15$; Pen_{base}: $r = -0.032$; PI_{base}: $r = 0.137$), because the base-binder effect is masked by the much larger variance caused by the SBS content. Controlling for the SBS content, the partial correlations are substantial: SP_{base}: $r = 0.54$, Pen_{base}: $r = -0.37$, and High PG_{base}: $r = 0.57$ (all $p < 0.001$), whereas PI_{base} remains weak ($r = 0.14$, $p = 0.21$). The base

binder therefore sets the reference level from which the polymer raises the grade, and its effect is material for design (Section 6.2). Pen_{base} and SP_{base} are moderately correlated (r = −0.61; Fig. 4); the large VIFs in Section 3.3 arise mainly because PI is a function of Pen and SP. The marginal correlation of the base-binder grade with the High PG of the modified binder is also weak (r = 0.029), although its partial correlation is 0.57.

Table 6. Pearson (r) and Spearman (ρ) correlation coefficients between the variables and the High PG.

Variable	n	Pearson r	Spearman ρ
SBS (%)	86	0.799 ***	0.796 ***
Pen_{base} (dmm)	85	−0.032 n.s.	−0.116 n.s.
SP_{base} (°C)	85	0.156 n.s.	0.208 n.s.
PI_{base}	85	0.137 n.s.	0.143 n.s.
Pen_{mod} (dmm)	56	−0.755 ***	−0.783 ***
SP_{mod} (°C)	71	0.831 ***	0.856 ***
PI_{mod}	56	0.615 ***	0.514 ***
High PG_{base} (°C)	84	0.029 n.s.	0.107 n.s.

Significance: *** p < 0.001; ** p < 0.01; * p < 0.05; n.s. = not significant. Pairwise complete records.

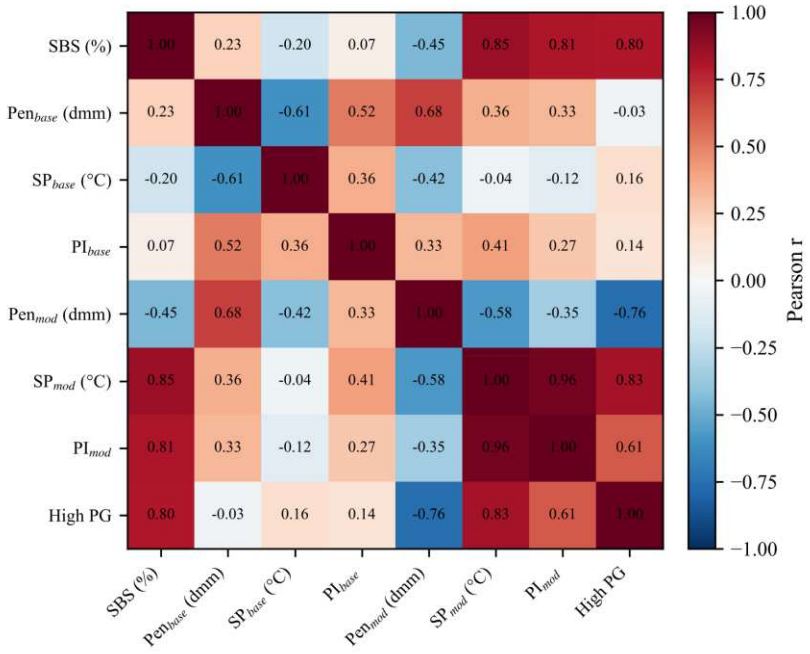


Fig. 4. Pearson correlation matrix of the variables of the High PG data set (pairwise complete records).

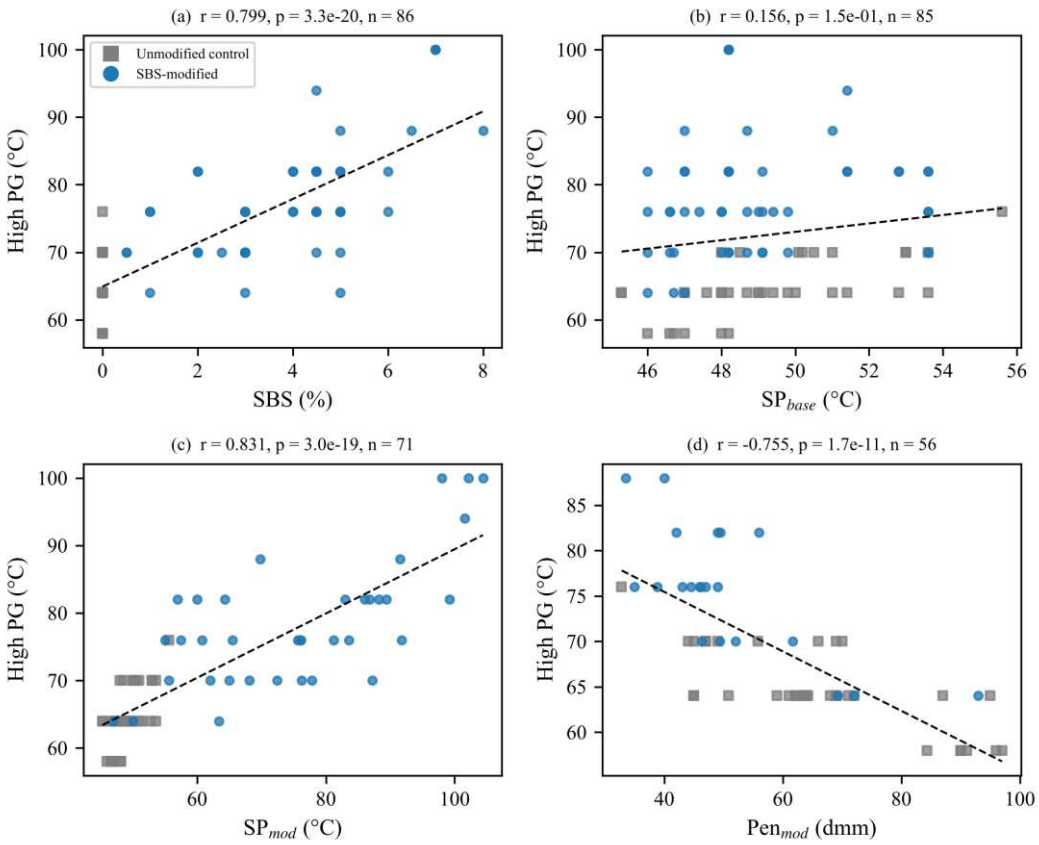


Fig. 5. Relationships between the High PG and (a) SBS content, (b) SP_{base}, (c) SP_{mod} and (d) Pen_{mod}, with least-squares lines, Pearson coefficients and p-values. Squares: unmodified controls; circles: SBS-modified binders.

3.3. Multiple linear regression and multicollinearity

MLR models fitted to all records for each scenario are summarized in Table 7. In Scenario H-I ($R^2 = 0.752$), SBS content is highly significant ($p < 0.001$, also with standard errors clustered by study [87]) and nearly independent of the other inputs (VIF = 1.1). By contrast, Pen_{base}, SP_{base}, and PI_{base} are collinear (VIF = 64–88) because PI is a function of Pen and SP, and none is significant on its own at $p < 0.05$ with ordinary standard errors (SP_{base}: $p = 0.069$); with cluster-robust standard errors, SP_{base} is significant ($p = 0.004$; Table 7). The SBS coefficient (3.61 °C per 1 % SBS; cluster-robust standard error 0.53 °C) is a pooled average over the controls and modified binders of all studies and is compared with other estimates of the SBS effect below. In Scenario H-II ($R^2 = 0.793$), SBS ($p < 0.001$) and Pen_{mod} ($p = 0.007$) are significant, whereas SP_{mod} and PI_{mod} are collinear (VIF ≈ 59 and 47) and not significant. In Scenario H-III ($R^2 = 0.893$, SE = 2.64 °C), SBS, Pen_{mod}, SP_{mod}, and PI_{mod} are significant ($p < 0.001$), whereas none of the base-binder terms is. The VIFs of SP_{mod} and PI_{mod} (≈ 88) indicate that the individual coefficients in Scenarios H-II and H-III should not be interpreted physically; the equations remain valid for prediction within the data range.

Table 7. MLR regression statistics for the three scenarios (all records of each scenario); p (cluster) = p-value with standard errors clustered by study.

Scenario	Term	Coefficient	Std. error	t	p	VIF	p (cluster)
H-I (R^2 = 0.752, n = 85)	Intercept	−137.917	119.021	−1.16	0.250	–	0.078
	SBS	3.610	0.237	15.22	< 0.001	1.1	< 0.001
	Pen_{base}	0.318	0.284	1.12	0.265	88.1	0.121
	SP_{base}	3.532	1.918	1.84	0.069	74.9	0.004
	PI_{base}	−9.143	7.471	−1.22	0.225	64.2	0.093
H-II (R^2 = 0.793, n = 56)	Intercept	66.626	21.479	3.10	0.003	–	0.022
	SBS	1.793	0.405	4.43	< 0.001	3.2	< 0.001
	Pen_{mod}	−0.188	0.067	−2.79	0.007	6.0	0.013
	SP_{mod}	0.198	0.326	0.61	0.546	58.9	0.660
	PI_{mod}	−0.947	1.839	−0.52	0.609	46.9	0.727
H-III (R^2 = 0.893, n = 55)	Intercept	−18.907	75.521	−0.25	0.803	–	0.787
	SBS	2.215	0.322	6.89	< 0.001	3.5	< 0.001
	Pen_{base}	0.258	0.175	1.47	0.148	63.3	0.099
	SP_{base}	0.341	1.212	0.28	0.779	65.7	0.763
	PI_{base}	5.144	4.855	1.06	0.295	45.1	0.254
	Pen_{mod}	−0.254	0.064	−3.99	< 0.001	9.3	< 0.001
	SP_{mod}	1.286	0.297	4.33	< 0.001	87.6	< 0.001
	PI_{mod}	−9.175	1.872	−4.90	< 0.001	87.9	< 0.001

Because records within a study share the base binder, SBS product, and laboratory, the ordinary standard errors in Table 7 are optimistic, and the pooled coefficient mixes within-study and between-study information. Table 8 therefore compares several estimates of the effect of 1 % SBS. The pooled MLR (3.61 °C per 1 %), the within-study estimate with study fixed effects (3.85 °C), and a mixed model with a random study intercept (3.67 °C) agree closely, and the median gain of the modified binders is 4.0 °C per 1 %. The rate is, however, not constant: it is larger for soft base binders (fixed effects for Pen_{base} > 75 dmm: 4.74 °C) than for 60/70 binders (3.42 °C; median 3.0 °C), it falls to 3.21 °C without the high-content study [13], and a mixed model with a quadratic term gives about 3.0 °C per 1 % between 0 and 3 % SBS. The lower rates obtained from the GPR design chart for a 60/70 binder (2.5 °C per 1 % up to 3 % SBS and 2.0 °C between 3 and 4.5 %; Section 6.2) and from the nomograph equations (1.6 and 2.2 °C in Eqs. (10) and (11)) are not contradictory: the nomograph coefficients are conditional on the Pen or SP of the blend, which already carries much of the SBS effect, and the GPR describes a specific base binder in the dosage range where most data lie. Applied to a 60/70 binder (65.6 °C), the pooled rate would give PG 76 at about 2.9 % SBS, whereas the 60/70 binders of the database reached PG 76 at 3–6 % SBS (median about 4.5 %), in agreement with the GPR chart (4.4 %). The GPR chart is therefore the design value for dosage selection, whereas the pooled 3.6 °C per 1 % is a database average that should not be used to compute dosages.

Table 8. Estimates of the effect of 1 % SBS on the High PG.

Estimate	Data / conditioning	°C per 1 % SBS (SE)
Pooled MLR, Scenario H-I (Table 7)	85 records; adjusted for Pen_{base}, SP_{base}, PI_{base}	3.61 (0.24; cluster-robust 0.53)
Linear regression on SBS content only	85 records	3.27 (0.27)
Study fixed effects (within-study)	85 records, 18 studies	3.85 (0.65)

Estimate	Data / conditioning	°C per 1 % SBS (SE)
Mixed model, random study intercept	adjusted for Pen_{base}, SP_{base}	3.67 (0.24)
Mixed model with SBS^2 term	slope 0–3 % SBS / at 4.5 % SBS	3.0 / 4.5
Median gain per 1 % SBS	49 modified binders with known base grade	4.0
Study fixed effects without study S12	67 records, 17 studies	3.21
Study fixed effects, 60/70 base binders	subset	3.42 (median 3.0)
Study fixed effects, Pen_{base} > 75 dmm	subset	4.74
GPR design chart, 60/70 binder (Section 6.2)	0–3 % / 3–4.5 % SBS	2.5 / 2.0
Eq. (10), nomograph	conditional on Pen_{mod} and PI_{base}	1.64
Eq. (11), nomograph	conditional on SP_{mod} and SP_{base}	2.18

SE = standard error. Rows 1–9 from statistical models of the database; rows 10–12 from the design chart and the nomograph equations. The design value for dosage selection is the GPR chart (Section 6.2).

4. Prediction results

Throughout this section, "test" refers to the 20 % hold-out set, "CV" to the nested 5 × 5-fold repeated cross-validation, and "LOSO" to the nested leave-one-study-out validation. All CV and LOSO statistics in this section were obtained with hyperparameters re-tuned within each outer fold (Section 2.5). Because the hold-out sets are small, a single test R^2 is a noisy statistic, and comparisons are based mainly on CV and LOSO. The performance of all models is listed in Table 9; the nested-CV and LOSO predictions of the best-ranked model of each scenario are plotted in Fig. 6, and those of all eight models are shown in Supplementary Figs. S1–S3.

4.1. High PG from the SBS content and the consistency of the base binder (Scenario H-I)

Using the SBS content and the Pen, SP, and PI of the base binder, all models explained 72–85 % of the variance in High PG in CV (Table 9). GPR was the most accurate (CV $R^2 = 0.845 \pm 0.011$, RMSE = 3.77 °C, 96 % within one grade), followed by XGBoost (0.831), ANN (0.819), LightGBM (0.812), RF (0.805), SVR (0.782); DT (0.728) and MLR (0.720) were the weakest. The no-change baseline gave $R^2 = -1.04$ (RMSE = 13.7 °C); the stronger baselines and the accuracy for the SBS-modified binders alone are compared in Section 4.4. In nested LOSO, the R^2 of GPR, MLR, RF, and XGBoost fell to 0.624–0.642, whereas LightGBM, ANN, SVR, and DT lost more (0.36–0.58); although MLR was among the least accurate models in CV, it transferred to unseen studies as well as the best non-linear model. A LOSO RMSE of about 5.7 °C means that a binder of a given penetration grade with a given SBS content can still differ by about one grade between laboratories, reflecting differences in base-binder chemistry, SBS product, and blending that are not captured by Pen and SP.

4.2. High PG from the consistency of the modified binder (Scenario H-II)

When the base-binder consistency was replaced with the modified-binder consistency, CV RMSE values were similar to those obtained with Scenario H-I (4.1 to 4.6 °C across all models except ANN and DT), but R^2 was lower because the modified-binder data set (H-II, 56 records) had fewer records and a smaller range of grades (standard deviation 7.5 °C, versus 9.6 °C in

H-I). On its own records, H-II (best CV RMSE 4.08 °C) was therefore not more accurate than H-I (3.77 °C), although it uses measurements of the blend: the H-II records contain only 21 SBS-modified binders, against 50 in H-I, and the modified-binder consistency of the controls equals that of the base binder, so H-II carries much less information on the polymer effect. On the 55 records common to all scenarios, H-II was more accurate than H-I (best CV RMSE 4.00 against 4.49 °C; Section 4.4). MLR had the highest CV R^2 (0.701, RMSE = 4.08 °C, 98 % within one grade), followed by GPR (0.676) and XGBoost (0.661); RF, LightGBM, and SVR reached 0.61–0.65, and DT (0.52) and ANN (0.46) were the least accurate. MLR and SVR transferred best to unseen studies (LOSO R^2 = 0.70 and 0.65; 100 % of the LOSO predictions of MLR within one grade), and ANN the least (LOSO R^2 = 0.49). The no-change baseline gave $R^2 = -0.31$ on the 54 records with a known base grade.

4.3. High PG from the consistency before and after modification (Scenario H-III)

Combining consistency before and after modification yielded the most accurate and transferable models. MLR and SVR (0.841 and 0.825; RMSE = 2.98 and 3.12 °C) had the highest CV accuracy, followed by GPR (0.803) and ANN (0.796). All four models had 100 % of predictions within one grade, and MLR had 69 % in the exact grade. In nested LOSO, MLR transferred best (LOSO R^2 = 0.85, RMSE = 2.88 °C), followed by ANN (0.77), GPR (0.75), and SVR (0.72), with 96–100 % of LOSO predictions within one grade. The polymer-induced change in Pen and SP thus contains information about the polymer network that can be transferred between labs. The tree-based models were evidently less precise in this case (CV R^2 = 0.52–0.66), since, with seven inputs and 55 records, their piecewise-constant approximations could not track the almost linear relationship between the grade and the consistency following the change.

Table 9. Performance of the eight models in the three High PG scenarios.

Scenario	Model	Train R^2	Test R^2	Test RMSE	CV R^2	CV RMSE	CV MAE	Exact (%)	±1 grade (%)	LOSO R^2	LOSO RMSE
H-I	MLR	0.748	0.746	5.46	0.720 ± 0.014	5.07	3.84	48	92.2	0.641	5.74
	SVR	0.972	0.862	4.02	0.782 ± 0.028	4.47	3.07	64	94.6	0.445	7.13
	GPR	0.992	0.860	4.06	0.845 ± 0.011	3.77	2.79	66	96.5	0.642	5.73
	DT	0.917	0.787	5.01	0.728 ± 0.026	4.99	3.76	49	92.9	0.359	7.67
	RF	0.970	0.839	4.35	0.805 ± 0.023	4.22	3.23	56	95.5	0.632	5.81
	XGBoost	0.995	0.807	4.77	0.831 ± 0.008	3.93	2.83	61	96.7	0.624	5.87
	LightGBM	0.977	0.794	4.92	0.812 ± 0.019	4.15	3.16	58	96.5	0.579	6.21
	ANN	0.911	0.835	4.41	0.819 ± 0.028	4.06	3.03	63	95.1	0.507	6.73
	Baseline (84)	–	–	–	−1.039	13.75	8.93	44	56	–	–
H-II	MLR	0.811	0.685	3.67	0.701 ± 0.016	4.08	3.19	57	97.9	0.701	4.09
	SVR	0.789	0.663	3.80	0.610 ± 0.072	4.65	3.37	56	94.6	0.651	4.41
	GPR	0.860	0.694	3.62	0.676 ± 0.025	4.25	3.35	55	96.1	0.632	4.53
	DT	0.916	0.751	3.26	0.518 ± 0.044	5.18	3.97	48	90.0	0.508	5.24
	RF	0.962	0.740	3.33	0.649 ± 0.013	4.42	3.53	49	94.6	0.597	4.74
	XGBoost	0.974	0.527	4.50	0.661 ± 0.032	4.34	3.41	54	94.6	0.613	4.65
	LightGBM	0.915	0.735	3.37	0.643 ± 0.075	4.44	3.44	52	95.0	0.603	4.71
	ANN	0.923	0.596	4.16	0.456 ± 0.188	5.44	3.77	56	92.5	0.489	5.34
	Baseline (54)	–	–	–	−0.312	8.56	4.67	65	74	–	–
H-III	MLR	0.916	0.759	3.28	0.841 ± 0.009	2.98	2.42	69	100.0	0.851	2.88
	SVR	0.941	0.734	3.45	0.825 ± 0.015	3.12	2.47	69	99.6	0.723	3.93
	GPR	0.929	0.751	3.33	0.803 ± 0.013	3.32	2.75	63	100.0	0.748	3.75

Scenario	Model	Train R^2	Test R^2	Test RMSE	CV R^2	CV RMSE	CV MAE	Exact (%)	±1 grade (%)	LOSO R^2	LOSO RMSE
	DT	0.827	0.533	4.56	0.520 ± 0.057	5.17	3.82	53	90.5	0.542	5.06
	RF	0.966	0.650	3.95	0.649 ± 0.031	4.42	3.50	49	94.5	0.632	4.54
	XGBoost	0.939	0.625	4.09	0.654 ± 0.052	4.39	3.32	54	92.4	0.545	5.04
	LightGBM	0.898	0.704	3.63	0.663 ± 0.028	4.33	3.42	51	93.5	0.554	4.99
	ANN	0.953	0.767	3.22	0.796 ± 0.009	3.38	2.76	60	99.6	0.774	3.55
	Baseline (54)	–	–	–	−0.312	8.56	4.67	65	74	–	–

RMSE and MAE in °C. CV = nested 5 × 5-fold repeated cross-validation (mean ± SD over five repetitions); LOSO = nested leave-one-study-out validation. Exact and ±1 grade accuracies refer to the repeat-averaged out-of-fold predictions rounded to the nearest standard grade. Baseline = no-change prediction $PG_{mod} = PG_{base}$, evaluated on the records with a known base grade (number of records in brackets; the same 54 records in H-II and H-III). The baseline involves no CV; its values are given in the CV columns for comparison only. The ±1 grade accuracy is given to one decimal, as used in the ranking of Section 4.4.

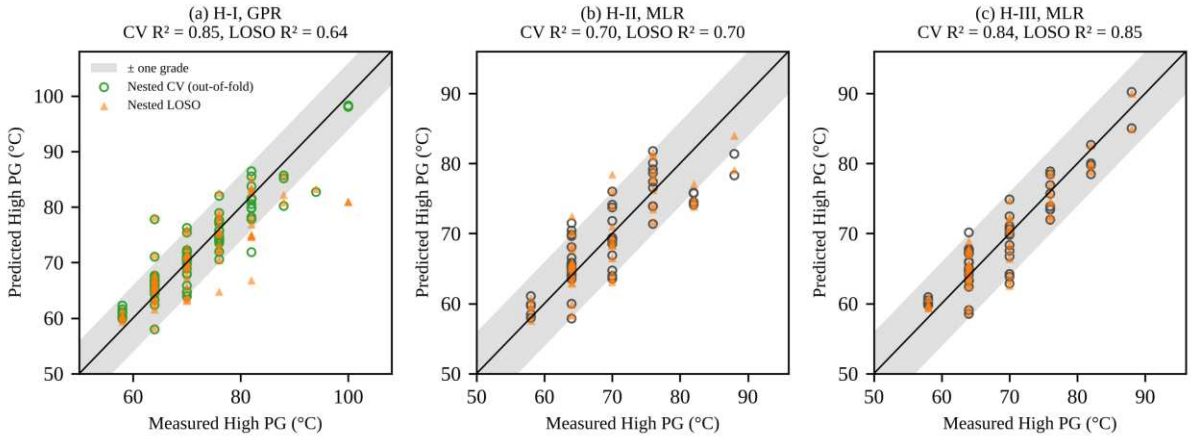


Fig. 6. Predicted versus measured High PG of the best-ranked model in each scenario. Circles: repeat-averaged out-of-fold predictions of the nested 5 × 5-fold CV; triangles: nested LOSO predictions; solid line: line of equality; shaded band: ± one grade (6 °C).

4.4. Comparison of models and scenarios

Fig. 7 compares the CV R^2, RMSE, and MAE for all 24 model-scenario combinations; Fig. 8 compares their distributions across the five CV replications; and Fig. 9 shows the Taylor diagrams of the out-of-fold predictions. Three patterns emerge. First, GPR and the boosted trees performed best when only the consistency of the base binder (H-I) was known, and differences among the leading models were small in all cases, with the lowest CV R^2 obtained by MLR in Scenario H-I, ANN in Scenario H-II, and DT in Scenario H-III (the lowest composite ranks were those of DT, ANN, and DT). Second, the linear MLR was the best model when the consistency of the modified binder was an input (H-II and H-III). This is consistent with an approximately linear dependence of the High PG on the polymer content and on the Pen and SP of the blend within the data range; however, with 55–56 records, of which only 20–21 are SBS-modified, and up to seven inputs, the bias–variance trade-off also favors simple models, so the advantage of MLR may reflect the sample size as much as the physics. Third, all models slightly underestimate the spread of the measured grades (points within the reference arc in Fig. 9), and the correlation between the out-of-fold predictions and the measurements for the highest-ranking models is 0.84–0.92.

Table 10 separates the accuracy for the SBS-modified binders from that for all records and compares the models with three practitioner baselines. Because the controls are predicted easily, R^2 over all records partly rewards separating modified from unmodified binders; for the SBS-modified binders alone the R^2 values are lower, and the RMSE in °C is the more transferable measure. For the modified binders, the best H-I models reached a CV RMSE of about 4.1 °C but a LOSO RMSE of 6.9–7.1 °C, i.e., about one grade, whereas the H-III MLR reached a LOSO RMSE of 2.7 °C (LOSO R^2 = 0.86) for the 20 modified binders of that scenario. The base grade plus an SBS slope fitted in the training folds ($PG_{mod} = PG_{base} + b \cdot SBS$) is a strong baseline over all records (LOSO R^2 = 0.62 on the H-I records and 0.80 on the H-II/H-III records) because it is exact for the controls; for the modified binders, however, its LOSO RMSE was 7.8 °C and 5.7 °C, compared with 7.1 °C for the H-I GPR and 2.7 °C for the H-III MLR. This baseline also requires the DSR grade of the base binder, which the models of Scenarios H-I to H-III do not. The SBS-only linear model was the weakest fitted reference (LOSO RMSE 6.5 °C over all H-I records, against 5.7 °C for the H-I GPR), so the Pen and SP of the base binder add information, and the Pen and SP of the blend (H-III) more than halve the error for the modified binders.

Table 10. Accuracy over all records and for the SBS-modified binders only, compared with practitioner baselines (nested CV and nested LOSO).

Scenario / model	n (all / mod.)	CV R^2	CV RMSE	LOSO R^2	LOSO RMSE	CV R^2 (mod.)	CV RMSE (mod.)	LOSO R^2 (mod.)	LOSO RMSE (mod.)
H-I GPR	85 / 50	0.85	3.77	0.64	5.73	0.77	4.09	0.33	7.05
H-I MLR	85 / 50	0.72	5.07	0.64	5.74	0.52	5.94	0.35	6.91
H-I XGBoost	85 / 50	0.83	3.93	0.62	5.87	0.77	4.15	0.31	7.14
H-I $PG_{base} + b \cdot SBS$	84 / 49	0.74	4.94	0.62	5.96	0.44	6.46	0.19	7.80
H-I SBS only (linear)	85 / 50	0.62	5.91	0.55	6.45	0.38	6.75	0.26	7.38
H-I No change	84 / 49	−1.04	13.75	–	–	−3.31	18.00	–	–
H-II MLR	56 / 21	0.70	4.08	0.70	4.09	0.42	5.26	0.41	5.31
H-II GPR	56 / 21	0.68	4.25	0.63	4.53	0.32	5.68	0.20	6.16
H-II $PG_{base} + b \cdot SBS$	54 / 19	0.81	3.30	0.80	3.36	0.41	5.56	0.39	5.67
H-II SBS only (linear)	56 / 21	0.53	5.13	0.52	5.18	0.21	6.12	0.25	5.97
H-II No change	54 / 19	−0.31	8.56	–	–	−2.96	14.44	–	–
H-III MLR	55 / 20	0.84	2.98	0.85	2.88	0.84	2.82	0.86	2.66
H-III SVR	55 / 20	0.83	3.12	0.72	3.93	0.83	2.91	0.43	5.33
H-III GPR	55 / 20	0.80	3.32	0.75	3.75	0.73	3.65	0.54	4.80
H-III ANN	55 / 20	0.80	3.38	0.77	3.55	0.85	2.78	0.68	4.03
H-III $PG_{base} + b \cdot SBS$	54 / 19	0.81	3.30	0.80	3.36	0.41	5.56	0.39	5.67
H-III SBS only (linear)	55 / 20	0.52	5.19	0.51	5.23	0.21	6.29	0.25	6.13
H-III No change	54 / 19	−0.31	8.56	–	–	−2.96	14.44	–	–

RMSE in °C. mod. = SBS-modified binders only (out-of-fold predictions of the same nested CV and LOSO runs, restricted to SBS > 0). $PG_{base} + b \cdot SBS$: slope b fitted within the training folds (records with a known base grade); SBS only: linear regression on the SBS content; No change: $PG_{mod} = PG_{base}$ (no fitting, hence no LOSO).

On the 55 records common to all scenarios (Table 11), the differences between the scenarios reflect only their inputs. The best H-I model reached a CV RMSE of 4.49 °C (MLR) and a LOSO R^2 of 0.66 (MLR), the best H-II model 4.00 °C (GPR) and 0.71 (MLR), and the best H-III model 3.02 °C (MLR) and 0.85 (MLR). On identical records, the consistency of the blend

is therefore more informative than that of the base binder, and combining both (H-III) gives the most accurate and transferable models, which confirms the ranking of the scenarios obtained on their own records (the H-III records are the common subset itself, so its values agree with Table 9 apart from the number of CV repetitions). The H-I R^2 values on this subset are lower than on all 85 H-I records because the subset has a smaller range of grades and contains only 20 SBS-modified binders. Ranking the models on the LOSO RMSE alone, instead of the composite of Table 12, gives the same leading models on the full scenario data sets: GPR, MLR, and RF in H-I; MLR, SVR, and GPR in H-II; and MLR, ANN, and GPR in H-III.

Table 11. Performance of the eight models in the three scenarios on the 55 records common to all scenarios (18 studies, 20 SBS-modified binders).

Model	H-I CV R^2	H-I CV RMSE	H-I LOSO R^2	H-I LOSO RMSE	H-II CV R^2	H-II CV RMSE	H-II LOSO R^2	H-II LOSO RMSE	H-III CV R^2	H-III CV RMSE	H-III LOSO R^2	H-III LOSO RMSE
MLR	0.64	4.49	0.66	4.39	0.71	4.00	0.71	4.05	0.84	3.02	0.85	2.88
SVR	0.54	5.07	0.60	4.75	0.69	4.15	0.64	4.47	0.82	3.20	0.72	3.93
GPR	0.58	4.83	0.55	5.04	0.71	4.00	0.66	4.35	0.80	3.34	0.75	3.75
DT	0.52	5.16	0.34	6.09	0.54	5.04	0.54	5.09	0.51	5.20	0.54	5.06
RF	0.57	4.88	0.54	5.08	0.62	4.60	0.61	4.69	0.63	4.54	0.63	4.54
XGBoost	0.54	5.06	0.38	5.89	0.65	4.40	0.63	4.54	0.64	4.50	0.54	5.04
LightGBM	0.51	5.25	0.52	5.17	0.66	4.35	0.63	4.53	0.64	4.47	0.55	4.99
ANN	0.46	5.46	0.35	6.02	0.70	4.09	0.56	4.97	0.80	3.34	0.77	3.55

RMSE in °C. CV = nested 3 × 5-fold repeated cross-validation (mean over three repetitions); LOSO = nested leave-one-study-out validation; hyperparameters re-tuned in every outer fold as in Section 2.5.

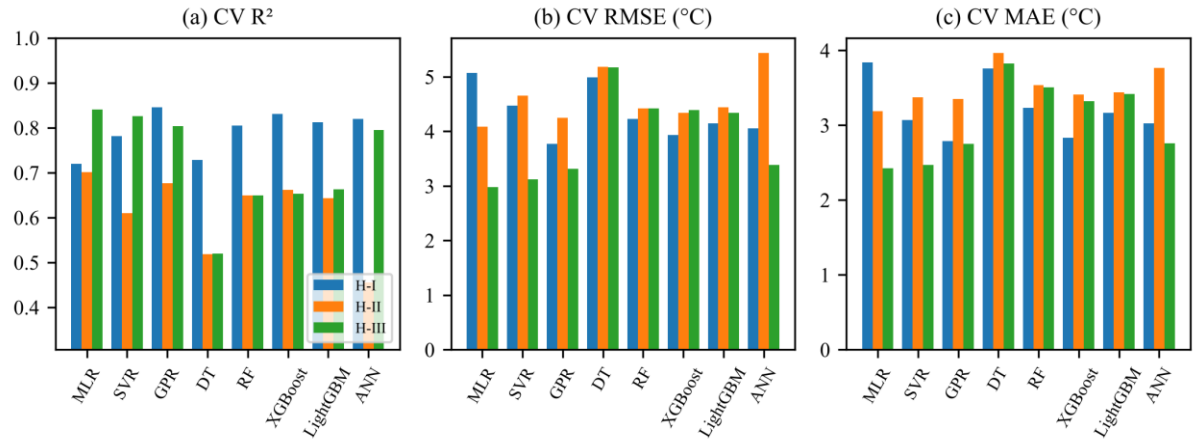


Fig. 7. Comparison of (a) cross-validated R^2, (b) RMSE and (c) MAE of the eight models in the three scenarios.

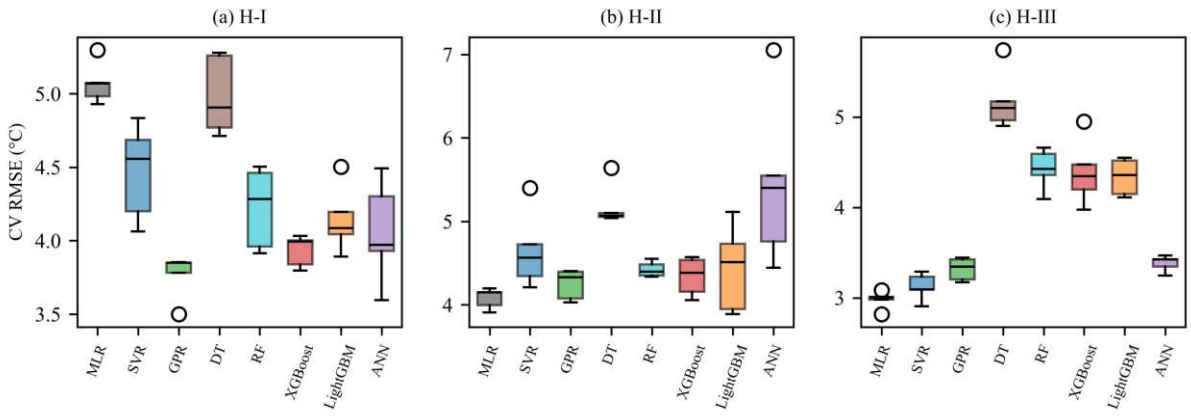


Fig. 8. Distribution of the cross-validated RMSE of the eight models over five repetitions of nested 5-fold cross-validation in each scenario.

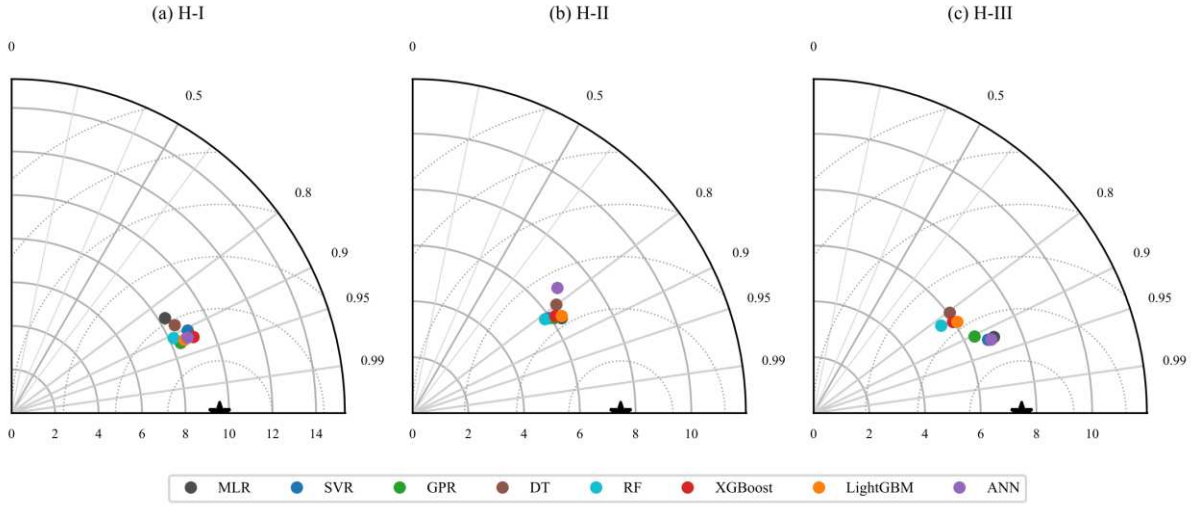


Fig. 9. Taylor diagrams of the out-of-fold predictions of the eight models in the three scenarios. The star represents the measured data; dotted arcs indicate the centered RMS difference.

The composite ranking (Table 12) placed GPR first in H-I, followed by XGBoost; MLR first in H-II, followed by GPR; and MLR first in H-III, followed by SVR and GPR (tied). The within-one-grade criterion uses the CV values before rounding (e.g., 99.6 % for SVR and ANN in H-III; Table 9), and ties receive the minimum rank. The Wilcoxon tests (Table 13) compare the absolute out-of-fold errors of the most accurate model in each scenario with those of the other models: in Scenario H-I, GPR had significantly smaller errors than MLR ($p < 0.001$; Holm-adjusted $p = 0.004$) and DT ($p = 0.003$; adjusted $p = 0.021$), but not than SVR, RF, XGBoost, LightGBM, and ANN; in Scenario H-II, MLR had smaller errors than DT ($p = 0.034$), but the difference was not significant after Holm's correction (adjusted $p = 0.24$), and no other difference was significant; and in Scenario H-III, MLR had significantly smaller errors than DT ($p = 0.002$; adjusted $p = 0.012$) and RF ($p = 0.004$; adjusted $p = 0.023$), whereas the difference from LightGBM ($p = 0.021$) was not significant after correction (adjusted $p = 0.11$),

and those from SVR, GPR, XGBoost, and ANN were not significant. Because the records of a study are not independent, these p-values are indicative. Based on accuracy, transferability, stability, and the availability of a predictive variance, GPR was used to derive the prediction formula and the design chart; it was ranked 1 in Scenario H-I, 2 in Scenario H-II, and 2 in Scenario H-III (ties have the same rank), and its CV R^2 was within 0.04 of the best model.

Table 12. Total ranking score (sum of the ranks on CV R^2, CV RMSE, CV MAE, LOSO R^2 and CV within-one-grade accuracy; ties receive the minimum rank) and final rank (in brackets; 1 = best) of the models in each scenario.

Model	H-I	H-II	H-III
MLR	34 (7)	5 (1)	5 (1)
SVR	29 (6)	22 (4)	13 (2)
GPR	6 (1)	11 (2)	13 (2)
DT	36 (8)	37 (7)	40 (8)
RF	23 (5)	25 (6)	31 (6)
XGBoost	11 (2)	18 (3)	31 (6)
LightGBM	20 (3)	23 (5)	28 (5)
ANN	20 (3)	38 (8)	17 (4)

Table 13. p-values of the Wilcoxon signed-rank test comparing the absolute out-of-fold errors of the most accurate model of each scenario (marked –) with those of the other models; Holm-adjusted p-values for the seven comparisons of each scenario in brackets.

Model	H-I (GPR)	H-II (MLR)	H-III (MLR)
MLR	< 0.001 (0.004)	–	–
SVR	0.815 (1.000)	0.757 (1.000)	0.887 (1.000)
GPR	–	0.161 (0.803)	0.169 (0.508)
DT	0.003 (0.021)	0.034 (0.238)	0.002 (0.012)
RF	0.121 (0.486)	0.123 (0.739)	0.004 (0.023)
XGBoost	0.079 (0.393)	0.282 (1.000)	0.073 (0.292)
LightGBM	0.835 (1.000)	0.504 (1.000)	0.021 (0.106)
ANN	0.904 (1.000)	0.257 (1.000)	0.657 (1.000)

4.5. Transferability to unseen studies

Fig. 10 compares nested CV and nested LOSO validation. The loss in R^2 when whole studies were omitted was approximately 0.2 for the best model in Scenario H-I, small in H-II (0.001 for MLR), and none in H-III, where the LOSO R^2 of MLR (0.851) was just as large as its CV R^2 (0.841). A LOSO R^2 above the CV R^2 is unusual but plausible here: each LOSO model is trained on about 94 % of the records (all but one study), against 80 % in 5-fold CV, and the pooled LOSO R^2 benefits from the between-study range of grades. Fig. 11 shows the error for each study when it was left out. The median (interquartile range) of the per-study LOSO RMSE was 3.5 (2.5–4.3) °C for the H-I GPR, 3.3 (1.3–5.1) °C for the H-II MLR, and 2.5 (1.6–3.4) °C for the H-III MLR. In Scenario H-III (Fig. 11b), the average error for each study was no more than ±4 °C. In Scenario H-I (Fig. 11a), the greatest errors were on an Iraqi study where a soft 80/100 binder was used and the binder was increased by up to four grades with 4 % or less SBS (mean error = +6.7 °C, i.e., under-predicted); a study of SBS contents up to 7 % and radial SBS (+4.2 °C); and a study of Turkey refinery binders, where 5 % SBS added to binders in two

refineries increased the grade by one to three grades, depending on the refinery and the penetration grade (RMSE = 5.3 °C). These instances indicate that the reaction of a given base binder to SBS cannot be determined completely from its Pen and SP unless a measurement of the blend is made. The most likely missing variables are the chemistry of the base bitumen (SARA fractions or colloidal instability index), the molecular weight and the styrene and vinyl contents of the SBS, and the blending conditions (temperature, shear rate, and time). Compared with a LOSO in which the hyperparameters of Table 4 were kept fixed, re-tuning within the folds had minor effects on the selected models; the LOSO R^2 decreased noticeably only for DT in Scenario H-I (by 0.15), SVR in Scenario H-I (by 0.14), and SVR in Scenario H-III (by 0.09). The reported accuracy is therefore not an artifact of tuning on the validation data.

The robustness of these results to the composition of the database was tested by repeating the nested CV and LOSO of the leading models after removing (i) the 29 grades reported by the authors (keeping the 57 records with full M 320 grading), (ii) the 5 figure-derived records, (iii) the four binders that gained five grades or more, and (iv) the largest study (S12, 18 records), and (v) after the correction of the base-binder penetration of one study (Table 14). The conclusions were stable. Restricting the data to full M 320 grades lowered the LOSO R^2 of the H-III MLR from 0.85 to 0.75 (RMSE 3.8 °C) on 45 records from 13 studies, whereas the H-I models were as accurate as before or better (LOSO R^2 = 0.67–0.72). Removing the binders with gains of five grades or more reduced the H-I LOSO RMSE to 4.1–4.7 °C. Removing study S12 lowered the LOSO R^2 of the H-I GPR to 0.52 but raised that of the H-I MLR to 0.70, so the advantage of GPR in H-I depends partly on the high-content data of this study. The penetration correction changed the R^2 values by at most 0.012.

Table 14. Sensitivity of the nested CV and LOSO accuracy of the leading models to the composition of the database.

Data set	Model	n (studies)	CV R^2	CV RMSE	LOSO R^2	LOSO RMSE	LOSO ±1 grade (%)
Full data set (as Table 9)	H-I GPR	85 (18)	0.85	3.74	0.64	5.73	91.8
	H-I MLR	85 (18)	0.73	4.99	0.64	5.74	91.8
	H-II MLR	56 (19)	0.70	4.11	0.70	4.09	100.0
	H-III MLR	55 (18)	0.84	3.02	0.85	2.88	100.0
	H-III GPR	55 (18)	0.80	3.34	0.75	3.75	96.4
Full M 320 grades only	H-I GPR	57 (13)	0.67	4.32	0.67	4.31	94.7
	H-I MLR	57 (13)	0.71	4.06	0.72	3.97	94.7
	H-II MLR	45 (13)	0.64	4.56	0.64	4.53	97.8
	H-III MLR	45 (13)	0.84	2.98	0.75	3.78	97.8
	H-III GPR	45 (13)	0.78	3.55	0.68	4.27	97.8
Figure-derived records removed (5)	H-I GPR	80 (16)	0.81	4.19	0.63	5.82	91.2
	H-I MLR	80 (16)	0.70	5.25	0.61	5.97	91.2

Data set	Model	n (studies)	CV R^2	CV RMSE	LOSO R^2	LOSO RMSE	LOSO ±1 grade (%)
	H-II MLR	54 (17)	0.68	4.23	0.71	4.06	100.0
	H-III MLR	53 (16)	0.84	3.03	0.85	2.92	100.0
	H-III GPR	53 (16)	0.77	3.61	0.74	3.86	96.2
Gains ≥ 5 grades removed (4)	H-I GPR	81 (18)	0.73	4.01	0.64	4.71	96.3
	H-I MLR	81 (18)	0.73	4.03	0.73	4.08	97.5
	H-II MLR	56 (19)	0.70	4.11	0.70	4.09	100.0
	H-III MLR	55 (18)	0.84	3.02	0.85	2.88	100.0
	H-III GPR	55 (18)	0.80	3.34	0.75	3.75	96.4
Study S12 removed (18)	H-I GPR	67 (17)	0.62	4.67	0.52	5.29	91.0
	H-I MLR	67 (17)	0.69	4.25	0.70	4.20	95.5
	H-II MLR	53 (18)	0.65	4.31	0.67	4.23	100.0
	H-III MLR	52 (17)	0.83	3.04	0.84	2.97	100.0
	H-III GPR	52 (17)	0.75	3.64	0.73	3.83	96.2
S29 Pen_{base} corrected (58.0 dmm)	H-I GPR	85 (18)	0.85	3.71	0.64	5.71	91.8
	H-I MLR	85 (18)	0.73	4.99	0.64	5.74	91.8
	H-II MLR	56 (19)	0.70	4.09	0.71	4.05	100.0
	H-III MLR	55 (18)	0.85	2.91	0.86	2.80	100.0
	H-III GPR	55 (18)	0.81	3.24	0.76	3.64	96.4

RMSE in °C. CV = nested 5 × 5-fold CV; LOSO = nested leave-one-study-out validation; hyperparameters re-tuned within every outer fold.

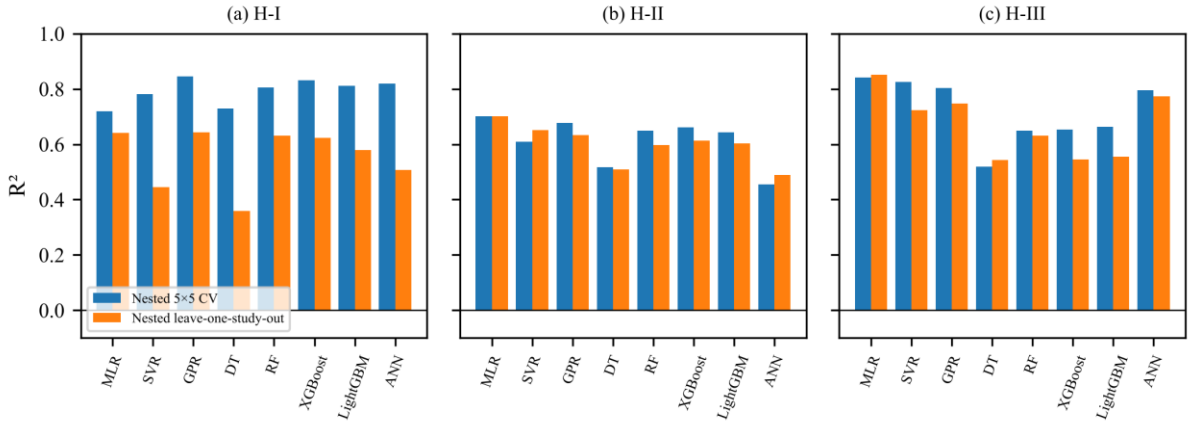


Fig. 10. Coefficient of determination obtained by nested 5 × 5-fold CV and by nested leave-one-study-out validation for each model and scenario.

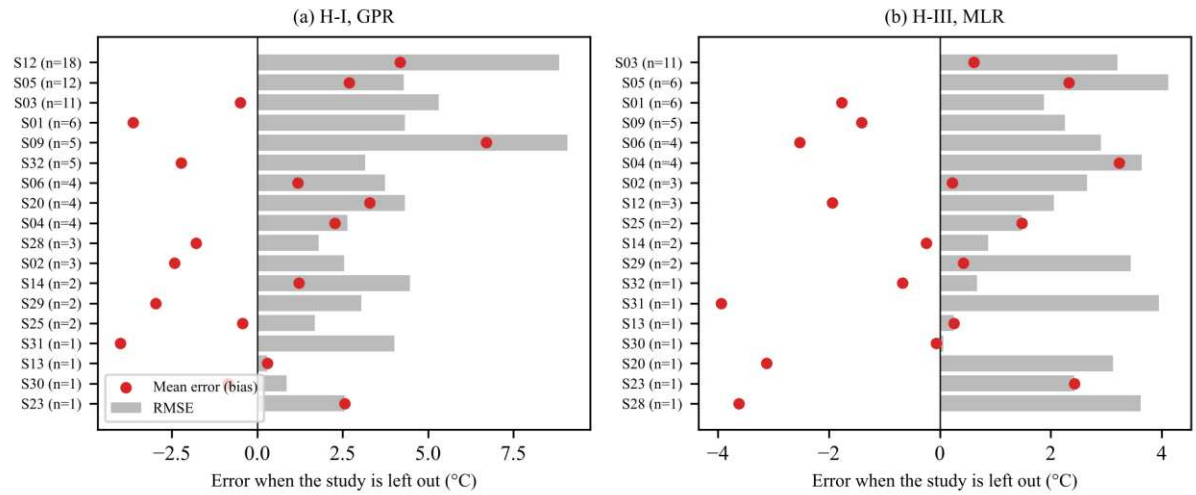


Fig. 11. Error of the best-ranked models of Scenarios (a) H-I (GPR) and (b) H-III (MLR) for each study when that study was left out of training: RMSE (bars) and mean error (measured − predicted; points). Studies are ordered by the number of records.

4.6. Residuals and grade-level accuracy

The out-of-fold residuals are centered on zero and mostly lie within ±6 °C (Fig. 12), with Scenario H-III showing the narrowest residuals. The confusion matrices of the best-ranked models (Fig. 13) show that misclassified grades are almost always off by one grade; in Scenario H-III, the MLR assigned 69 % of the binders to the exact grade and 100 % to the exact or an adjacent grade. Underprediction mainly concerns the binders with the highest grades (PG 88 and above in Scenario H-I and PG 82 and above in Scenario H-II), which lie at the edge of the data. For design, over-prediction (a predicted grade higher than the measured one) is the unsafe error. It occurred for 12.9 %, 19.6 %, and 12.7 % of the LOSO predictions of the best H-I, H-II, and H-III models (CV: 15.3 %, 21.4 %, and 14.5 %), and about half (H-I) to most (H-II and H-III) of the over-predicted binders had SBS contents of 3 % or less, where regression towards the mean raises the prediction.

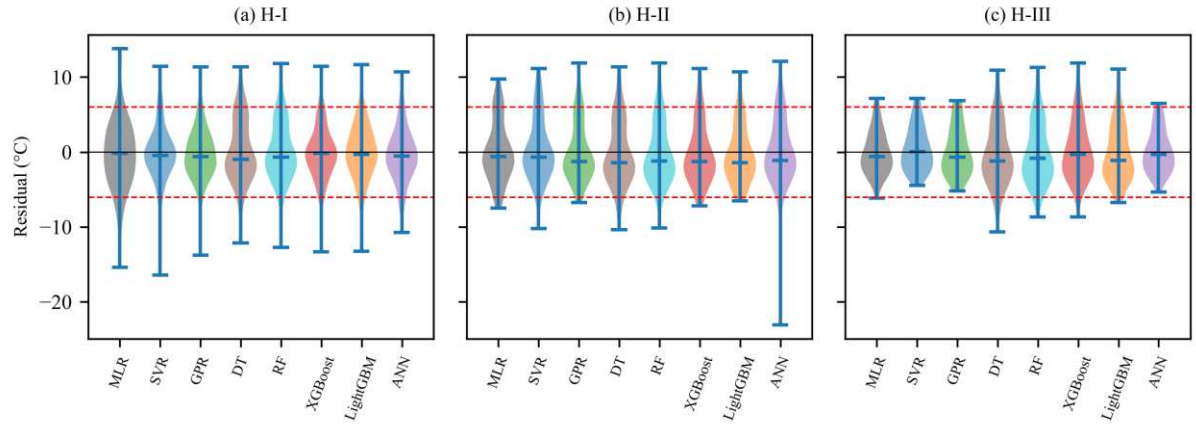


Fig. 12. Distribution of the out-of-fold residuals of the eight models in the three scenarios (dashed red lines: ± one grade increment of 6 °C).

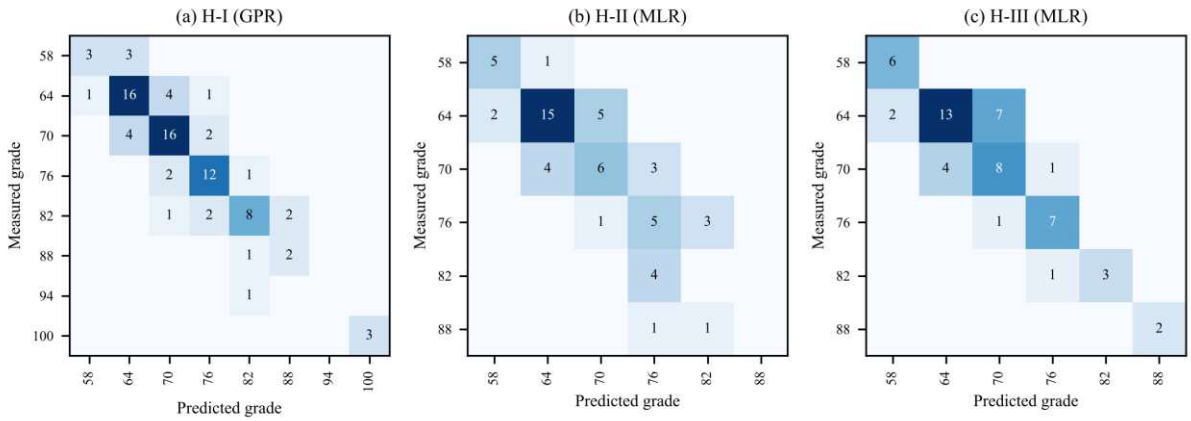


Fig. 13. Confusion matrices of the measured and predicted grades (repeat-averaged out-of-fold predictions rounded to the nearest standard grade) for the best-ranked model of each scenario.

5. Importance of the input variables

The importance of the inputs, shown as relative and absolute values, is indicated in Fig. 14 and Table 15, respectively. In Scenario H-I, the most influential input of SVR, GPR, DT, RF, XGBoost, LightGBM, and ANN (42–70 % of the total importance) was the SBS content; for MLR, the collinear SP_{base} received slightly more (8.20 °C, 34 %) than the SBS content (7.87 °C, 32 %); Pen_{base}, SP_{base}, and PI_{base} received the rest. In Scenario H-II, the weight shifted toward the consistency of the modified binder; the most influential input was Pen_{mod} for SVR, GPR, XGBoost, LightGBM, and ANN; SP_{mod} for DT and RF; and SBS for MLR, with SBS receiving 18–40 percent. In Scenario H-III, the modified-binder variables together contributed more than half of the total importance in all models except the ANN, which attributed large importances to Pen_{base} (5.25 °C) and SP_{base} (5.18 °C), so that the modified-binder variables account for only about 39 % of its total. The very large in-sample importances of SP_{mod} and PI_{mod} for MLR (18–21 °C) are an artifact of permuting functionally linked inputs, which creates impossible Pen–SP–PI combinations [84]; the grouped, out-of-fold analysis below avoids this.

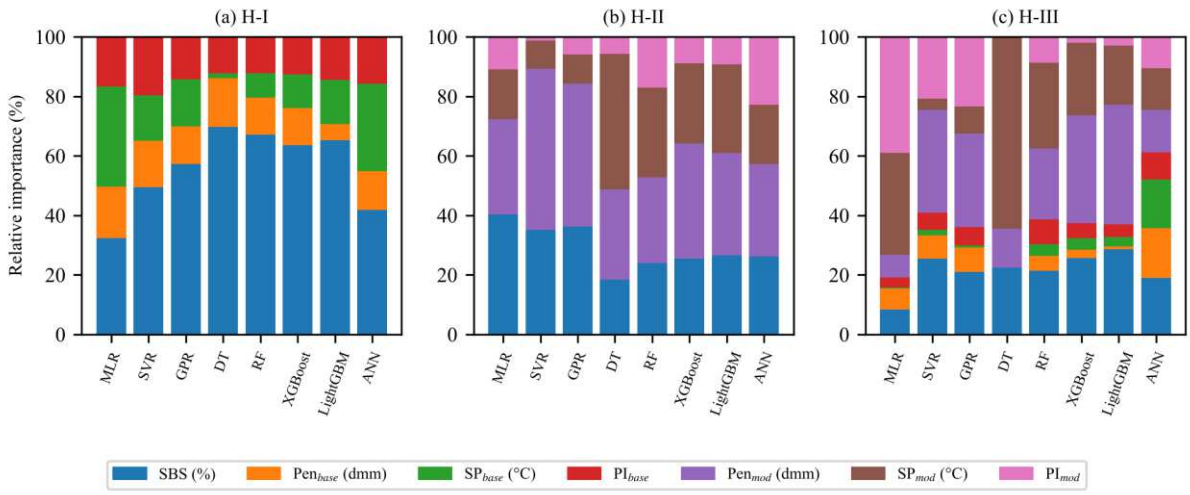


Fig. 14. Relative permutation importance (%) of the input variables for the eight models in each scenario.

Table 15. Permutation importance (mean increase in RMSE, °C, over 20 permutations) of the inputs in the three scenarios.

Scenario	Input	MLR	SVR	GPR	DT	RF	XGB	LGBM	ANN
H-I	SBS (%)	7.87	10.28	10.48	9.64	10.42	11.17	10.69	9.69
	Pen_{base} (dmm)	4.25	3.24	2.31	2.27	1.93	2.19	0.89	3.05
	SP_{base} (°C)	8.20	3.15	2.89	0.23	1.28	2.00	2.44	6.82
	PI_{base}	4.08	4.09	2.60	1.70	1.90	2.22	2.37	3.66
H-II	SBS (%)	2.81	1.99	2.41	2.05	2.12	2.38	2.40	3.16
	Pen_{mod} (dmm)	2.24	3.07	3.20	3.38	2.56	3.61	3.12	3.77
	SP_{mod} (°C)	1.16	0.54	0.66	5.06	2.67	2.52	2.71	2.43
	PI_{mod}	0.77	0.07	0.40	0.64	1.52	0.83	0.83	2.77
H-III	SBS (%)	4.48	5.34	4.54	1.54	2.10	2.31	2.37	5.92
	Pen_{base} (dmm)	3.72	1.65	1.76	0.00	0.51	0.24	0.07	5.25
	SP_{base} (°C)	0.19	0.37	0.15	0.00	0.38	0.36	0.26	5.18
	PI_{base}	1.82	1.23	1.34	0.00	0.83	0.45	0.36	2.87
	Pen_{mod} (dmm)	4.08	7.27	6.76	0.90	2.34	3.25	3.33	4.50
	SP_{mod} (°C)	18.27	0.77	1.99	4.43	2.85	2.18	1.64	4.38
	PI_{mod}	20.77	4.39	5.03	0.00	0.86	0.19	0.25	3.30

XGB = XGBoost; LGBM = LightGBM.

The SHAP summary plots for the best-ranked nonlinear model in each scenario are shown in Fig. 15, and the mean absolute SHAP values are presented in Table 16. In the GPR for Scenario H-I, SBS content has by far the largest mean |SHAP| (6.66 °C), followed by SP_{base} (1.71 °C). Its SHAP values rise monotonically from about −7 °C for unmodified binders to about +8 °C at 4–5 % and +17 °C at 6–8 % SBS, relative to the base value φ_0 = 70.6 °C, i.e., the mean prediction over the ten-centroid background (the mean measured grade is 72.5 °C). A high SP_{base} or a low Pen_{base} raises the prediction. In the SVR for Scenario H-III, Pen_{mod} has the largest mean |SHAP| (5.24 °C), followed by SBS (4.45 °C) and PI_{mod} (3.51 °C). A low Pen_{mod} and a high SBS content raise the predicted grade, whereas, at a given Pen_{mod} and SP_{mod}, a high PI_{mod} lowers it, which partly compensates for the collinear SP_{mod} contribution. The same directions were found for the GPR of Scenario H-II, in which Pen_{mod} (3.05 °C) and SBS (2.74 °C) dominate.

The grouped, out-of-fold permutation importances and the grouped SHAP values (Table 17) avoid impossible input combinations. In H-I, permuting the SBS content increased the out-of-fold RMSE by 7.6–8.7 °C, compared with 1.8–3.8 °C for the base-binder group; in H-II, the modified-binder group (2.6–4.7 °C) was more influential than the SBS content (0.4–2.4 °C) for all models, with similar values for MLR; and in H-III, the modified-binder group dominated (3.0–8.2 °C), followed by the SBS content (0.7–4.0 °C) and the base-binder group (0–4.8 °C). The grouped SHAP values give the same order (H-I GPR: SBS 6.66 °C, base group 3.19 °C; H-III SVR: modified group 4.87 °C, SBS 4.45 °C, base group 2.65 °C). With Pen, SP, and PI treated jointly, the conclusions on the relative roles of the SBS content and of the base- and modified-binder consistency are unchanged, whereas the individual shares of Pen, SP, and PI within a group (Table 15 and Fig. 14) should not be interpreted separately.

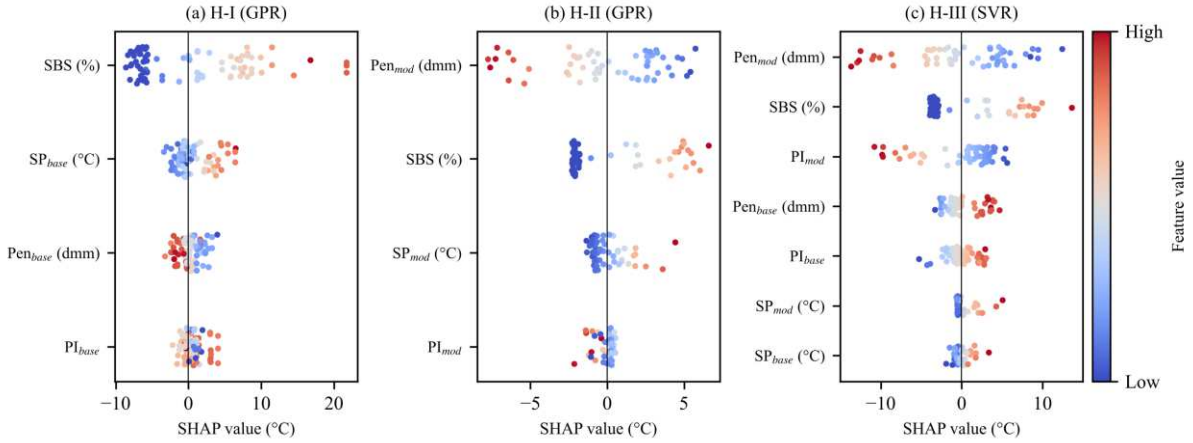


Fig. 15. SHAP summary plots of the best-ranked non-linear model in each scenario. Inputs are ordered by mean absolute SHAP value; color indicates the input value (blue = low, red = high).

Table 16. Mean absolute SHAP values (°C) of the input variables, in descending order.

Scenario	Model	Input	Mean |SHAP| (°C)
H-I	GPR	SBS (%)	6.66
		SP_{base} (°C)	1.71
		Pen_{base} (dmm)	1.10
		PI_{base}	1.01
H-II	GPR	Pen_{mod} (dmm)	3.05
		SBS (%)	2.74
		SP_{mod} (°C)	0.95
		PI_{mod}	0.43
H-III	SVR	Pen_{mod} (dmm)	5.24
		SBS (%)	4.45
		PI_{mod}	3.51
		Pen_{base} (dmm)	1.77
		PI_{base}	1.31
		SP_{mod} (°C)	0.74
		SP_{base} (°C)	0.66

Table 17. Grouped out-of-fold permutation importance (mean increase in RMSE, °C) and grouped mean |SHAP| values (°C).

Scenario	Input group	MLR	SVR	GPR	DT	RF	XGB	LGBM	ANN	SHAP
H-I	SBS content	7.55	8.09	8.40	7.83	7.92	8.70	8.63	8.36	6.66

Scenario	Input group	MLR	SVR	GPR	DT	RF	XGB	LGBM	ANN	SHAP
	Base binder {Pen, SP, PI}	1.78	3.83	3.11	2.09	2.22	2.85	3.06	2.93	3.19
H-II	SBS content	2.37	1.43	1.58	0.39	0.67	0.57	0.80	2.19	2.74
	Modified binder {Pen, SP, PI}	2.56	3.04	2.63	3.88	3.50	3.13	3.75	4.67	3.54
H-III	SBS content	3.77	4.00	3.19	0.69	0.77	0.95	1.06	2.92	4.45
	Base binder {Pen, SP, PI}	4.81	2.44	1.84	-0.01	0.23	0.15	0.15	1.74	2.65
	Modified binder {Pen, SP, PI}	8.18	5.50	5.21	3.63	3.15	3.22	2.98	4.29	4.87

Out-of-fold permutation importance: 3 × 5-fold CV, models refitted (with the hyperparameters of Table 4) on the training folds, inputs of a group permuted jointly (10 permutations) in the held-out fold. SHAP: sum of the SHAP values of the group for the best non-linear model (H-I and H-II: GPR; H-III: SVR).

6. Prediction formula and SBS design chart

6.1. GPR prediction formula

To enable the use of the models outside the original Python environment (e.g., in a spreadsheet), the GPR for each scenario was refitted to all its records using the tuned kernel. Table 18 lists the parameters, and the fitted weights and standardized reference inputs are provided in the supplementary spreadsheet; the formula reproduces the fitted models exactly and should be used only within the range of the training data. The inputs are first standardized as $z_j = (x_j - \mu_j)/\sigma_j$, and the predicted grade is the posterior mean:

$$\text{High PG} = \bar{y} + s_y \sum_i \alpha_i \, k(z, z_i), \quad r_i = \|z - z_i\| / \ell \tag{8}$$

$$k(z, z_i) = c^2(1 + \sqrt{3}\, r_i) \exp(-\sqrt{3}\, r_i) \quad (\nu = 1.5) \tag{9}$$

For $\nu = 0.5$ (Scenario H-I), the kernel reduces to $c^2 \exp(-r_i)$, and for $\nu = 2.5$ (Scenario H-III) it becomes $c^2(1 + \sqrt{5}\, r_i + 5r_i^2/3) \exp(-\sqrt{5}\, r_i)$. Here, z_i are the standardized inputs of the training records, α_i the fitted weights, ℓ the length scale, c^2 the signal variance, and $\bar{y}$ and s_y the mean and standard deviation of the measured grades. The MLR equations (Table 7) offer a simpler alternative. Both should be applied only within the database range (Table 2), i.e., for SBS contents up to 8 % and base binders of penetration grades 30/40 to 80/100 (Pen_{base} = 33–97 dmm). Eq. (8) is not a closed-form equation for hand calculation: it requires all 55–85 training vectors and weights α_i, which are provided in the supplementary spreadsheet so that it can be implemented in a spreadsheet. For H-III, the fitted length scale ($\ell \approx 20$ in the standardized space) and signal variance ($c^2 = 71.5$) make the kernel almost linear over the data, which is consistent with the good performance of MLR in this scenario.

Table 18. Parameters of the GPR prediction formulas (models refitted on all records of each scenario).

Scenario	n	ν	ℓ	c^2	σ_n^2	$\bar{y}$ (°C)	s_y (°C)	Standardization μ_j/σ_j (inputs in the order of Table 3)
H-I	85	0.5	22.738	5.281	0.078	72.47	9.58	2.32/2.34; 68.84/17.64; 49.14/2.41; −0.72/0.57
H-II	56	1.5	6.828	3.730	0.192	68.82	7.47	1.51/2.09; 60.21/17.23; 55.88/11.20; 0.28/1.77
H-III	55	2.5	20.028	71.498	0.121	68.69	7.48	1.45/2.06; 68.80/16.18; 48.98/2.38; −0.76/0.49; 60.67/17.04; 55.71/11.23; 0.26/1.78

c^2 and σ_n^2 refer to the standardized target (High PG − $\bar{y}$)/s_y; σ_j is the population standard deviation.

6.2. SBS design chart

The GPR for Scenario H-I, which requires only the SBS content and the Pen and SP of the base binder, was used to construct the design chart in Fig. 16 for three typical penetration-grade binders: 40/50 (Pen 45 dmm, SP 51 °C), 60/70 (Pen 65 dmm, SP 48 °C), and 80/100 (Pen 90 dmm, SP 46 °C). Because the tuned kernel of this scenario ($\nu = 0.5$) follows the scatter between neighboring records and yields locally non-monotonic curves, the chart was drawn using the same GPR refitted with a smoother Matérn kernel ($\nu = 1.5$); this yields curves that increase monotonically with the SBS content for all three binders (checked numerically on a 0.01 % SBS grid; minimum slopes 1.7, 1.8, and 0.04 °C per 1 % for the 40/50, 60/70, and 80/100 binders), at almost the same accuracy (CV $R^2 = 0.83$ and LOSO $R^2 = 0.60$, against 0.85 and 0.64). The 80/100 curve is nearly flat between 4 and 5 % SBS, between about 2.7 and 4.2 % SBS it lies slightly (< 1 °C) above the 60/70 curve, and beyond about 6.6 % SBS the 60/70 curve rises above the 40/50 curve; a harder base binder giving a lower grade at the same SBS content is physically implausible. These features reflect the sparse data at high SBS contents: only 4 of the 85 records (from two studies) exceed 6.5 % SBS, three of them from one study with PG 100. The chart was therefore truncated at 6.5 % SBS (dashed beyond this content in Fig. 16a), and differences of less than about 1 °C between the binder curves are not meaningful. For the 60/70 binder, the predicted mean High PG rises from 65.6 °C without SBS to 73.1 °C at 3 % SBS (2.5 °C per 1 %) and 76.2 °C at 4.5 % SBS (2.0 °C per 1 % between 3 and 4.5 %), so that mean dosages of about 1.7, 4.4, and 6.1 % SBS are needed to reach PG 70, PG 76, and PG 82 (Fig. 16b); PG 88 (7.5 %) lies in the extrapolated region.

The harder 40/50 binder already reaches PG 70 and requires about 3.1 % and 5.7 % SBS for PG 76 and PG 82, whereas the softer 80/100 binder requires about 2.1 % and 5.0 % for PG 70 and PG 76; PG 82 for the 80/100 binder (6.8 %) and PG 88 for all binders lie beyond the supported range. The shaded bands in Fig. 16a are 95 % predictive intervals, which include the noise variance σ_n^2 of the GPR: for the 60/70 binder, the interval is about ±6.9 °C at 3 % SBS, of which the noise term ($\sigma_n \approx 2.9$ °C) accounts for most; the 95 % credible interval of the mean curve is about ±3.8 °C. At 2.0–2.5 °C per 1 % SBS, the predictive interval corresponds to roughly ±3 % SBS, i.e., the dosage uncertainty for an individual binder is as large as the dosage itself, which reflects the between-study scatter discussed in Section 4.5. Because under-dosing is the unsafe error, Fig. 16b also gives a reliability-based dosage, defined as the SBS content at which the lower one-sided 80 % predictive bound ($\mu - 0.84\sigma$) reaches the target grade: for the 60/70 binder, about 2.9 % and 5.4 % SBS for PG 70 and PG 76 (5.9 % for PG 76 at 90 % reliability), and for the 40/50 binder, 2.2 % and 5.5 %; the 80 %-reliability dosage for PG 76

with the 80/100 binder (6.6 %) and for PG 82 with any binder exceed the supported range. To check whether the smoother kernel affects the dosages, the required SBS contents were also computed with ν = 0.5 (the tuned kernel) and ν = 2.5. Across the 9 binder–grade combinations reached by the design chart (excluding PG 70 for the 40/50 binder, which the base binder already reaches) and the two alternative kernels, the results differed from those of the design chart by an average of 0.30 % SBS, and 14 of the 18 comparisons differed by 0.5 % SBS or less (PG 88 for the 40/50 and 80/100 binders is not reached with ν = 1.5 and ν = 2.5); the largest difference (1.5 % SBS) occurred for the 80/100 binder at PG 76, where the ν = 0.5 curve gives 3.6 % against 5.0 % for ν = 1.5. The choice of ν therefore has little effect on the dosages for most binders and grades, and where it does, the difference is smaller than the uncertainty band of the chart.

For hot-climate design, the chart indicates that typical Iraqi and Middle Eastern 40/50 and 60/70 binders require roughly 3.1–4.4 % SBS (mean) or 5.4–5.5 % SBS (80 % reliability) to reach PG 76, the grade required for most of the United Arab Emirates and much of Iraq [4,5], and about 5.7–6.1 % (mean) for PG 82, the grade that results from the bump of one or two grades for slow or heavy traffic in AASHTO M 323 [88]. Contents of about 6 % SBS approach the practical limits of storage stability and of the viscosity at mixing and compaction temperatures for many binders [10,18], so PG 82 may require a compatible base binder, a stabilized PMB, or another modification route. The chart should be used to select a starting dosage for laboratory verification, not to replace DSR testing.

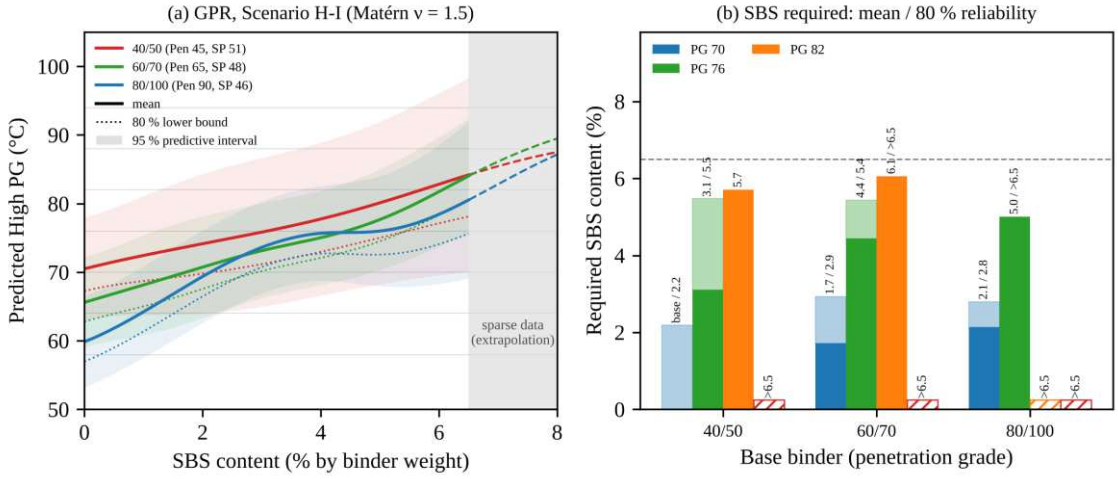


Fig. 16. GPR-based design chart (Scenario H-I, kernel ν = 1.5): (a) predicted High PG versus SBS content for 40/50, 60/70 and 80/100 base binders (solid lines: mean; dotted lines: lower one-sided 80 % predictive bound; shaded bands: 95 % predictive interval; dashed lines and gray area: SBS > 6.5 %, extrapolation); (b) SBS content required to reach PG 70, PG 76, PG 82 and PG 88: mean (solid bars) / 80 % reliability (light extension; values "mean / reliability") ("base": grade already reached by the base binder; ">6.5": not reached within the supported range of 6.5 % SBS).

7. Nomographs for practical High PG estimation

The prediction formula in Section 6 requires software, and the design chart in Fig. 16 covers only three typical base binders. This section presents three-input alignment nomographs and heat maps of the same equations that can be used without software once a trial blend has been prepared and its consistency measured, so the blend's High PG can be estimated before it is sent for DSR grading and the SBS content can be adjusted if necessary.

7.1. Selection of the inputs

The best of the 89 three-input models screened are listed in Table 19. Most of the models that transferred best to unseen studies included the penetration of the modified binder (logarithmic or linear), and the highest LOSO R^2 was obtained with PI_{base}, log Pen_{mod}, and the SBS content ($R^2 = 0.84$, LOSO $R^2 = 0.79$). Replacing PI_{base} with Pen_{base}, log Pen_{base}, or SP_{base} lowered the LOSO R^2 by at most 0.04. The model based on the softening point alone (SP_{base}, SP_{mod}, and SBS) was only slightly less accurate ($R^2 = 0.80$, LOSO $R^2 = 0.75$) and could be calibrated on 70 instead of 55 records, because the penetration of the modified binder is reported less often than its SP. In contrast, the combination of the penetration indices before and after modification with the SBS content ranked 86 of 89 ($R^2 = 0.57$, LOSO $R^2 = 0.42$): PI_{mod} is strongly correlated with the SBS content ($r = 0.81$), and the PI condenses Pen and SP into a single number that does not retain the absolute stiffness level on which the High PG depends; for the SBS binders, PI_{mod} mainly reflects the elevation of the SP by the polymer network rather than temperature susceptibility (Section 2.2). The most accurate model and the SP-based model were therefore presented as nomographs.

Because the candidate models were screened and selected using the same database, some optimism in the reported validation performance cannot be ruled out. To quantify this optimism, the entire screen-and-select procedure was repeated within an outer LOSO loop on the 55 binders (18 studies) with Pen and SP before and after modification: for each held-out study, all 89 candidate models were screened by LOSO on the remaining studies only, the best was refitted, and the held-out study was predicted. This selection-aware validation yielded LOSO $R^2 = 0.78$ (RMSE = 3.47 °C, 100 % within one grade), compared with 0.79 (RMSE = 3.40 °C) for the fixed PI_{base}–Pen_{mod}–SBS model on the same records, and the same input combination was selected in 17 of the 18 outer folds. The optimism caused by screening is therefore small. Nevertheless, the nomographs should be regarded as preliminary screening tools, and their wider applicability should be confirmed using independent experimental datasets.

Table 19. The ten best three-input MLR models for the High PG ranked by LOSO R^2, and the model based on the penetration indices.

Inputs	n (studies)	R^2	Adj. R^2	SE (°C)	CV R^2	LOSO R^2	LOSO RMSE (°C)
PI_{base}, log Pen_{mod}, SBS	**55 (18)**	**0.84**	**0.83**	**3.15**	**0.80**	**0.79**	**3.40**
PI_{base}, Pen_{mod}, SBS	55 (18)	0.83	0.82	3.21	0.79	0.79	3.45
SP_{base}, log Pen_{mod}, SBS	55 (18)	0.81	0.80	3.35	0.76	0.77	3.58
SP_{base}, Pen_{mod}, SBS	55 (18)	0.81	0.80	3.38	0.76	0.77	3.58
SP_{base}, SP_{mod}, SBS	**70 (18)**	**0.80**	**0.79**	**4.67**	**0.75**	**0.75**	**4.97**
log Pen_{base}, log Pen_{mod}, SBS	55 (18)	0.81	0.80	3.39	0.77	0.75	3.72
Pen_{base}, log Pen_{mod}, SBS	55 (18)	0.81	0.80	3.41	0.76	0.75	3.73
log Pen_{base}, Pen_{mod}, SBS	55 (18)	0.80	0.79	3.47	0.75	0.75	3.76
Pen_{base}, Pen_{mod}, SBS	55 (18)	0.80	0.79	3.46	0.75	0.75	3.77
log Pen_{base}, SP_{mod}, SBS	70 (18)	0.80	0.79	4.60	0.76	0.74	5.09
PI_{base}, PI_{mod}, SBS (rank 86)	55 (18)	0.57	0.55	5.07	0.42	0.42	5.68

Bold: models drawn as nomographs (Figs. 17 and 18). Each model was fitted on all records with the three inputs available; the complete list of 89 models is given in the supplementary results file.

7.2. Nomograph from the PI of the base binder, the penetration of the modified binder and the SBS content

Among the 89 screened combinations, the highest LOSO R^2 was obtained with the penetration index of the base binder (PI_{base}), the logarithm of the penetration of the modified binder (log Pen_{mod}), and the SBS content (Table 19). This combination corresponds to the verification stage, in which the base binder is characterized by its Pen and SP through the PI (Eq. (1)) and the trial blend by its measured penetration. The MLR equation fitted to all 55 records (18 studies) for which the three inputs were available is given in Eq. (10). Its statistical performance, its physical interpretation, and the corresponding alignment nomograph are presented below, and the calibration ranges within which it can be applied are stated at the end of this section.

$$\text{High PG} = 138.15 + 3.088\ PI_{base} - 39.34 \log Pen_{mod} + 1.637\ SBS \quad (10)$$

(n = 55, 18 studies; $R^2 = 0.84$; adjusted $R^2 = 0.83$; SE = 3.15 °C; all terms $p \leq 0.002$; VIF ≤ 1.52). The equation transferred to unseen studies (CV $R^2 = 0.80$; LOSO $R^2 = 0.79$, RMSE = 3.40 °C), and 100 percent of the LOSO predictions fell within a single grade, with 58 percent within the exact grade. The greatest proportion of information is in Pen_{mod}: removing it reduced R^2 to 0.57, compared with removing SBS content to 0.69 and removing PI_{base} to 0.80. The positive PI_{base} coefficient is consistent with expectations: at the same blend penetration and SBS content, a less temperature-susceptible base binder (PI_{base} refers to the unmodified binder, for which the PI is meaningful) retains more stiffness at the DSR test temperatures. The SBS term (1.6 °C per 1 % SBS at a given Pen_{mod}) represents the elastic contribution of the polymer network to $|G^*|/\sin\delta$, which is not reflected in the penetration at 25 °C.

Eq. (10) is shown in Fig. 17. A point on the turning line T is indicated by a line between PI_{base} and Pen_{mod}, and a second line from that point to the SBS content crosses the High PG scale.

The example is a base binder with PI = −0.70, 4 % SBS, and Pen_{mod} = 55 dmm, which is projected to reach 74.1 °C, i.e., PG 70. The nomograph is considered valid within the calibration ranges, i.e., PI_{base} = −2.2 to 0.3, Pen_{mod} = 33–97 dmm, and SBS content = 0–6.5 %. Because the inputs are correlated (high SBS contents go with a low Pen_{mod}), combinations inside the individual ranges can lie outside the data (e.g., Pen_{mod} = 90 dmm with 6 % SBS); the region supported by data is shown in the heat maps of Fig. 20.

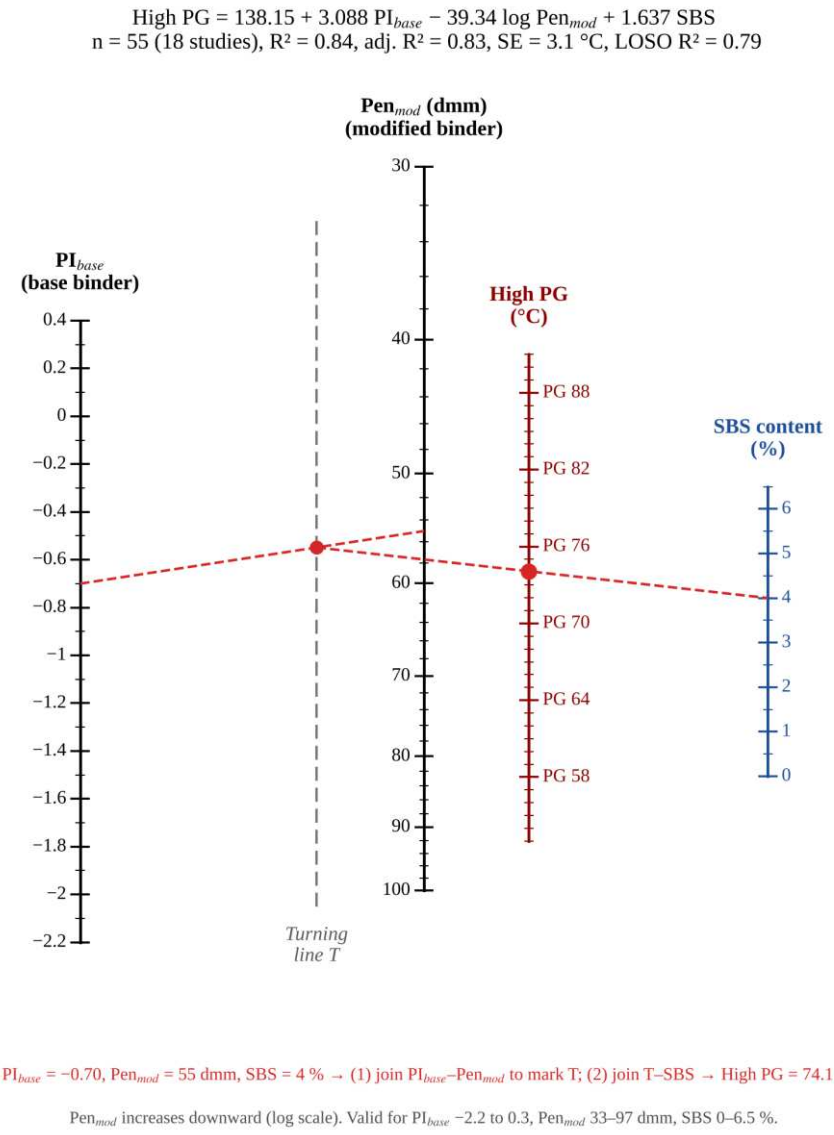


Fig. 17. Three-input alignment nomograph for the High PG from the penetration index of the base binder, the penetration of the modified binder (logarithmic scale) and the SBS content (Eq. (10)); the SBS scale is limited to the calibration range (0–6.5 %); the dashed line shows the worked example.

7.3. Nomograph from the softening point before and after modification and the SBS content

$$\text{High PG} = -1.94 + 1.113\ SP_{base} + 0.231\ SP_{mod} + 2.184\ \text{SBS} \qquad (11)$$

(n = 70, 18 studies; $R^2 = 0.80$; adjusted $R^2 = 0.79$; SE = 4.67 °C; all terms $p \leq 0.001$; VIF ≤ 4.0; CV $R^2 = 0.75$; LOSO $R^2 = 0.75$, RMSE = 4.97 °C; 93 % of LOSO predictions within one grade). The most informative term was the SBS content ($R^2 = 0.73$ without it), then SP_{base} (0.74), and then SP_{mod} (0.75). The base binder SP determines the grade reference level, and the SBS content contributes approximately 2.2 °C per 1 % increase over the SP of the blend, indicating that the ring-and-ball test underestimates the high-temperature rigidity caused by the polymer network. The SP_{mod} coefficient is small (0.23 °C per °C, against 1.11 for SP_{base}), so that, once the SBS content is known, SP_{mod} is a weak indicator of the grade, contrary to the common practice of judging a polymer-modified binder by its softening point; removing any one term lowers R^2 only from 0.80 to 0.73–0.75, which indicates strong redundancy among the three inputs. Eq. (11) is shown in Fig. 18; for a base binder with SP = 48 °C modified with 4 % SBS to SP_{mod} = 65 °C, the predicted High PG is 75.2 °C, i.e., PG 70. The nomograph can be used with SP_{base} between 45 and 56 °C, SP_{mod} between 45 and 104 °C, and SBS contents as high as 7 %. The precision of the input tests should be considered when the nomographs are read: a reproducibility of about 2 °C for the softening point (ASTM D36 [27]; larger for polymer-modified binders) propagates to about ±2.2 °C in the High PG through the SP_{base} term of Eq. (11) and ±0.5 °C through the SP_{mod} term, and an error of ±4 dmm in a penetration of 55 dmm propagates to about ±1.2 °C through Eq. (10). These errors account for a large part of the leeway between grade boundaries discussed in Section 7.4.

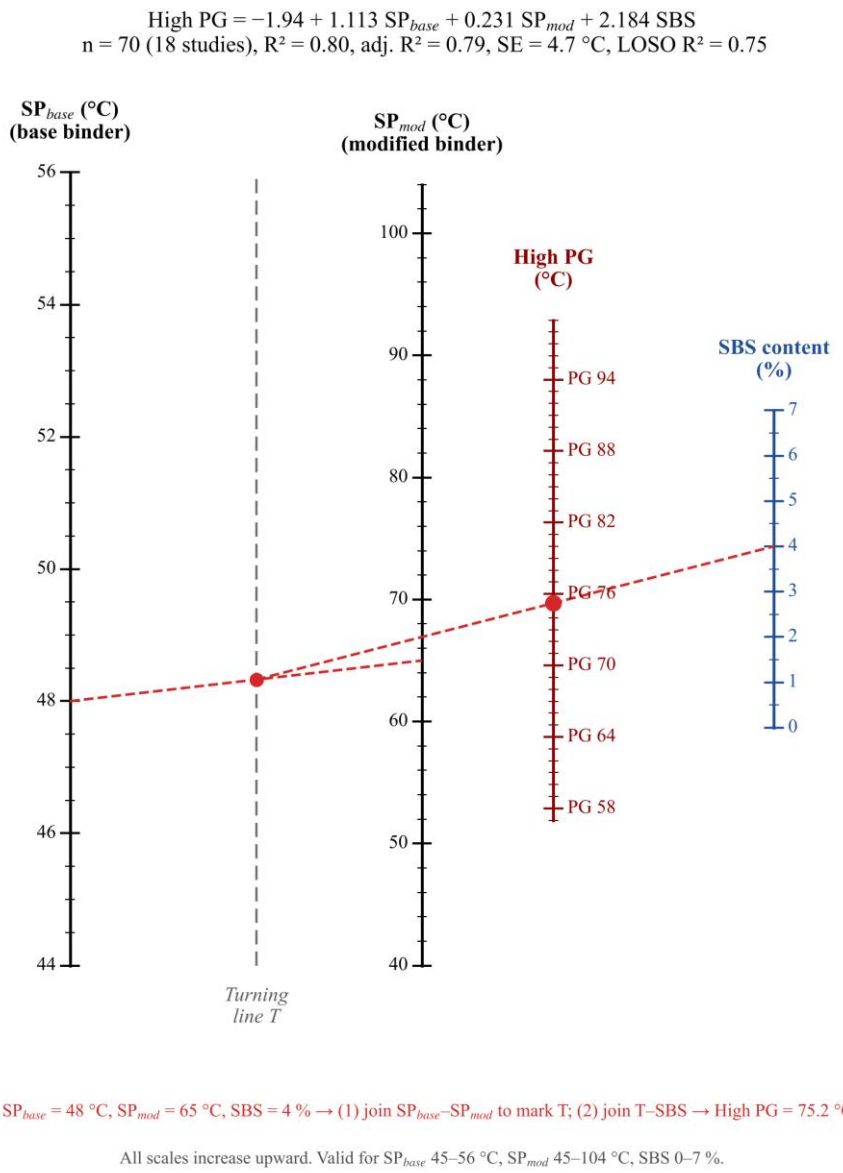


Fig. 18. Three-input alignment nomograph for the High PG from the softening points of the base and modified binders and the SBS content (Eq. (11)); the dashed line shows the worked example.

The fitted and LOSO predictions for both nomographs are compared in Fig. 19. Because SP is reported more frequently, the SP-based nomograph was fit to more records and covers a broader range of grades. However, on the 55 records for which both equations could be evaluated, it was less accurate (R^2 = 0.72, LOSO R^2 = 0.63, LOSO RMSE = 4.6 °C) than the nomograph in Fig. 17 (R^2 = 0.84, LOSO R^2 = 0.79, LOSO RMSE = 3.4 °C). When the blend is available, the more informative measure is the blend's penetration.

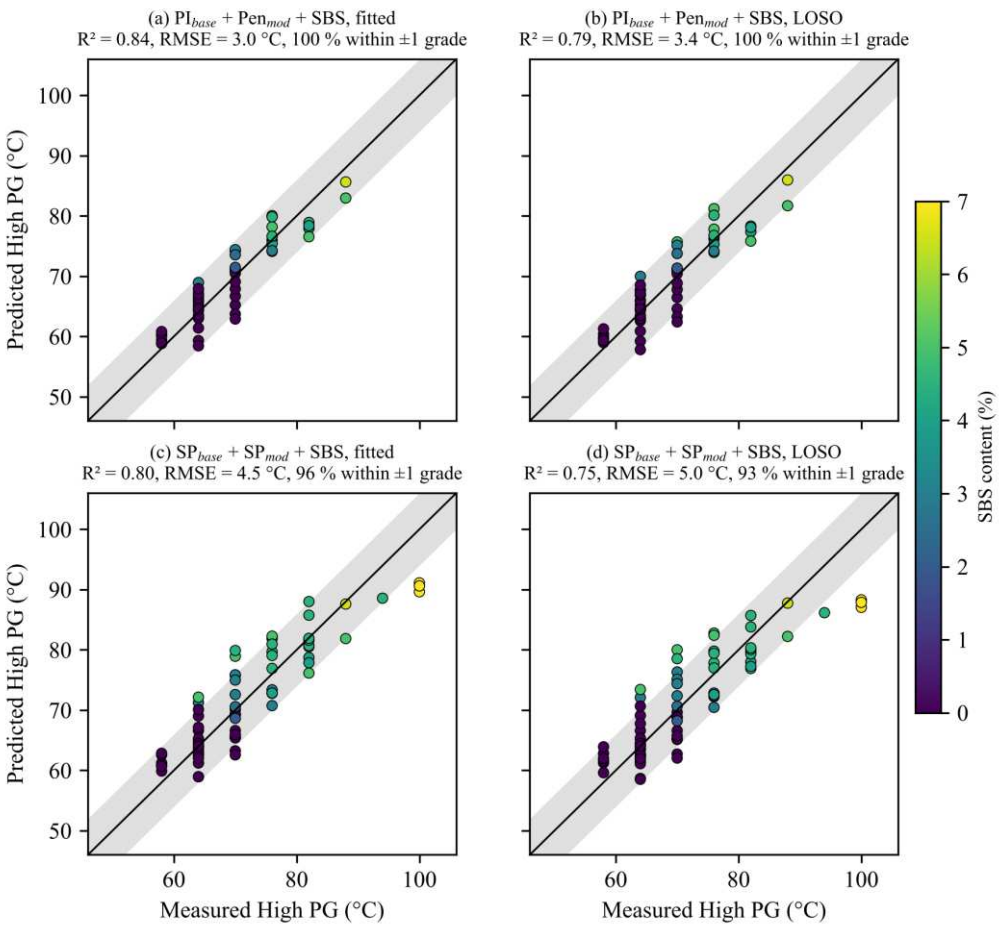


Fig. 19. Measured versus predicted High PG of the nomograph equations: (a, b) PI_{base}, Pen_{mod} and SBS content (Fig. 17), fitted and LOSO; (c, d) SP_{base}, SP_{mod} and SBS content (Fig. 18), fitted and LOSO. Grey band: ±6 °C (one grade).

7.4. Heat maps of the nomograph equations

A nomograph of alignment will give a single value for one combination of the inputs but will not indicate the gradient of the grade as the entire range of the consistency inputs is varied. The two equations were also plotted as two-dimensional heat maps of SBS contents of 2, 4, and 6 % (Fig. 20), where the color indicates the predicted grade band, the contour lines show the grade limits, and the measured binders with SBS contents within a range of ±1 % of the mapped value are superimposed. Regions without data support, identified from the nearest-neighbor distance to the calibration records in the standardized space of the three inputs, are masked (dotted), and predictions below PG 58, the lowest grade in the data, are grayed out. The grade boundaries of the maps of Eq. (10) (Fig. 20a–c) are curved, since the penetration enters logarithmically: at a particular PI_{base}, a reduction of Pen_{mod} from 90 to 60 dmm increases the High PG by 6.9 °C, whereas a further reduction from 60 to 30 dmm increases it by 11.8 °C. The change in all boundaries with a 1 % SBS shift is 1.6 °C, and thus, at a certain blend penetration, PI_{base}, and SBS content, a 2 % SBS shift causes a binder to shift by approximately half a grade. The boundaries in the maps of Eq. (11) (Fig. 20d–f) are straight lines with a slope

of −4.8 °C of SP_{mod} per 1 °C of SP_{base}: a 1 °C harder base binder is approximately 5 °C of SP_{mod}, and the area $SP_{mod} < SP_{base}$, not reachable by SBS, is hatched. Most measured binders lie in or next to the band of their measured grade, which illustrates the accuracy of about one grade reported in Sections 7.2 and 7.3. The heat maps are appropriate for selecting a target Pen_{mod} or SP_{mod} for a trial blend: given a base binder and an SBS content, the band to which the blend must go can be read directly, and the distance to the next boundary indicates how much leeway a measured blend has, to be judged against the test precision discussed in Section 7.3.

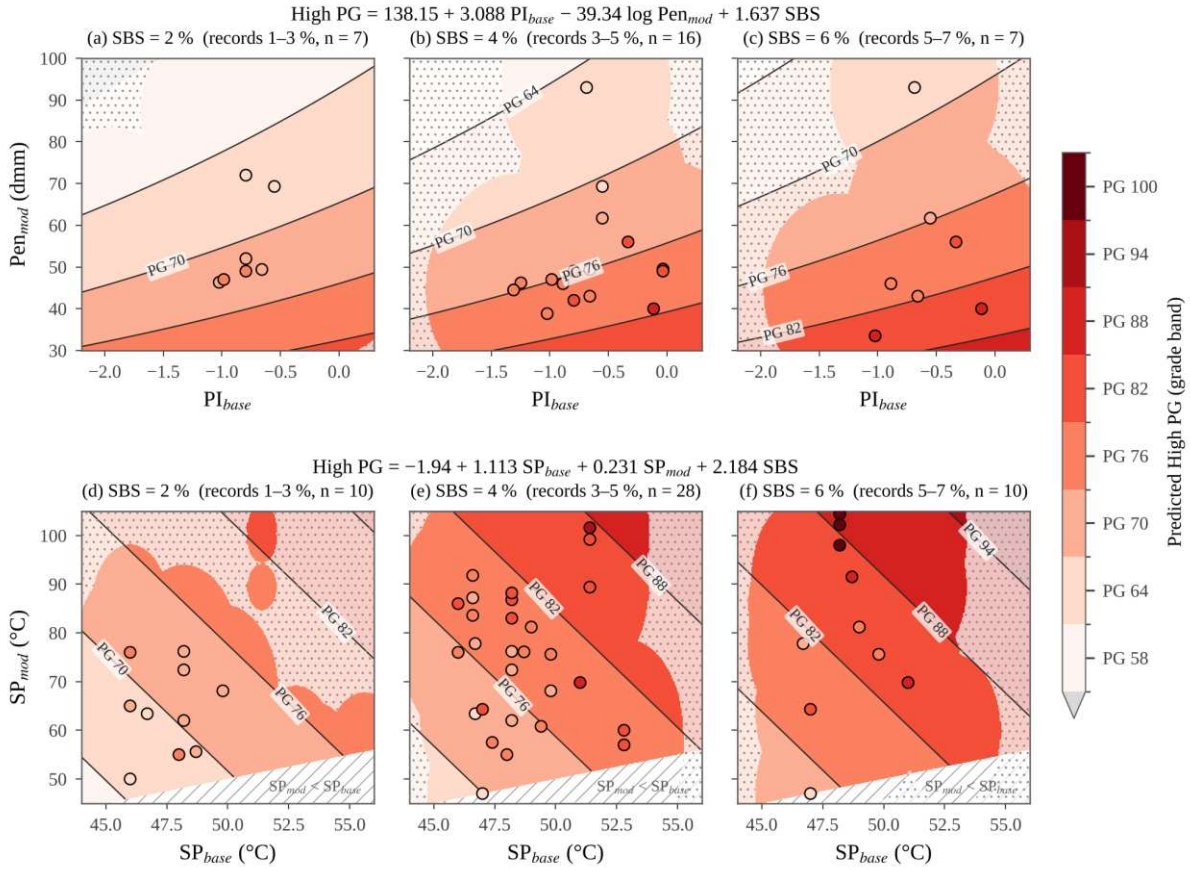


Fig. 20. Heat maps of the High PG predicted by the nomograph equations at SBS contents of 2, 4 and 6 %: (a–c) Eq. (10) over PI_{base} and Pen_{mod}; (d–f) Eq. (11) over SP_{base} and SP_{mod}. Color: predicted grade band (gray: below PG 58, outside the calibration range); dotted overlay: regions without data support; lines: grade boundaries; circles: measured binders with SBS contents within ±1 % of the mapped value, colored by their measured grade; hatched: $SP_{mod} < SP_{base}$.

7.5. Importance of polymer structure as an input parameter

Fig. 21a shows that the out-of-fold residuals for the H-I GPR differed by SBS type: radial SBS binders were under-predicted by a mean of 2.6 °C (n = 7), and linear SBS binders were over-predicted by 1.2 °C (n = 25) (Mann–Whitney p = 0.03). In the 8 paired comparisons of radial and linear SBS at the same base binder and content within the same study (Fig. 21b), radial SBS yielded a higher upper PG temperature (continuous failure temperature) in 7 cases, by a

mean of 2.8 °C (Wilcoxon p = 0.016), consistent with the stronger network of radial and higher-molecular-weight polymers [12,14]. The pairs come from three studies: five from [13], two from [14], and one from [89], whose failure temperatures were measured on the original binder only and are therefore not part of the High PG data set (hence 8 pairs but 7 radial binders with a High PG in Table 2). Because the pairs within a study are not independent, the p-value overstates the evidence; at the study level, the mean difference was positive in all three studies (3.4, 2.2, and 1.5 °C). The residual of the SP gain was also higher for radial SBS (Fig. 21c; Mann–Whitney p = 0.001), but this difference is confounded by the studies: radial SBS was used in only 4 studies in the database, of which 2 provide High PG data, and in 1 of them the radial polymers gave a lower SP than the linear ones at the same content.

The polymer structure was therefore not included as an input to the nomographs and heat maps: its apparent effect (about +3 °C in the paired comparisons) is smaller than the between-study scatter of the High PG predictions, it is based on 7 radial binders with a High PG from 2 studies, and too few studies report the architecture to calibrate a separate scale. Other SBS properties, such as the styrene content (20–33 %, mostly about 30 %, in the studies that reported it), the vinyl content, and the molecular weight, are reported too rarely to be modeled. The available results suggest that radial SBS may produce slightly higher High PG values than linear SBS under comparable conditions. However, the limited number of radial SBS records prevents reliable generalization of this difference.

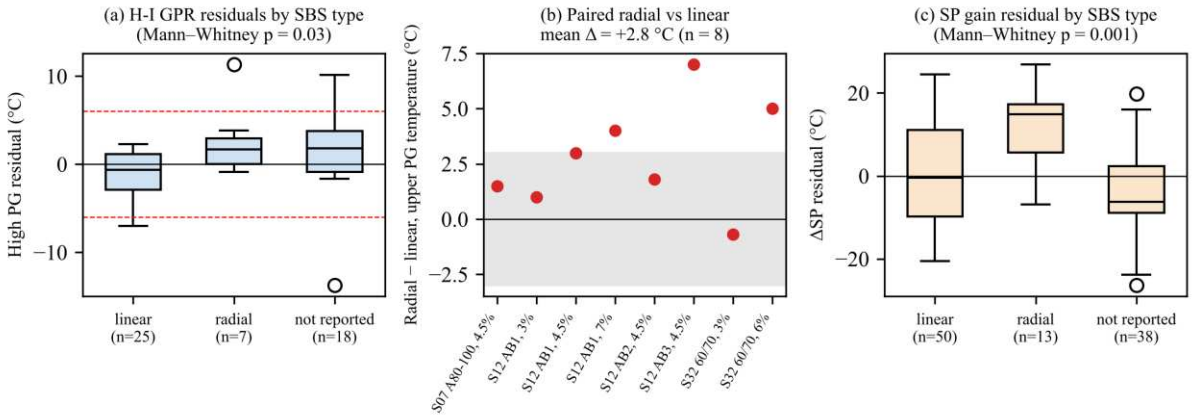


Fig. 21. Effect of the SBS architecture: (a) out-of-fold residuals of the H-I GPR by SBS type; (b) paired differences of the upper PG temperature between radial and linear SBS at the same base binder and SBS content in the same study; (c) residual of the SP gain (after SBS content) by SBS type.

7.6. Recommended use

The design chart, the nomographs, and the heat maps support a three-step process. (1) During the design phase, a starting SBS content is selected using either the design chart (Fig. 16; the reliability-based dosage where under-dosing must be avoided) with the base binder's Pen and

SP, or the GPR formula of Section 6 and the heat maps (Fig. 20) to determine the target Pen_{mod} or SP_{mod} that the trial blend must attain. (2) A trial blend is prepared, and its Pen and SP are measured. The nomograph of Fig. 17 (when the blend's penetration is measured) or of Fig. 18 (when only the softening point is measured) provides its High PG, and the SBS content is adjusted until the target grade is achieved. (3) The final binder is graded according to AASHTO M 320. Steps (1) and (2) reduce the number of trial blends that must be sent for DSR grading. The charts are not to be applied outside the ranges specified in Table 2 and in Sections 7.2 and 7.3, and in general not with base binders softer than 80/100. The proposed approach can inform laboratories about promising formulations and minimize unnecessary trial-and-error testing when developing a binder by providing preliminary estimates of the required SBS content and the desired binder grade.

8. Discussion

8.1. Mechanistic interpretation

The findings, based on a multi-study database, substantiate the well-known stiffening effect of SBS at high temperatures and quantify it in grade terms. Across the database, the High PG increased by about 3.3–3.9 °C per 1 % SBS (pooled, within-study, and mixed-model estimates; Table 8), but the rate depends on the base binder and decreases in the dosage range used for design, where the design chart gives 2.0–2.5 °C per 1 % for a 60/70 binder; the median modified binder (4.5 % SBS) gained 2 grades. This is greater than would be suggested by the change in SP alone: at a constant SP_{mod}, every 1 % SBS will increase the elastic contribution to |G*| and decrease the phase angle at the DSR test temperatures [90,91], which is only partially reproduced in the ring-and-ball test. The analysis of importance reveals that, after the blend has been prepared, the majority of the information about the grade is carried by its Pen and SP, and the base-binder consistency is important as a reference level. This is consistent with the swelling mechanism: the degree to which the polymer constitutes a continuous phase varies with the maltene level and the compatibility of the base bitumen [16,17], which vary between refineries of identical penetration grade and are indicated by the consistency of the blend but not by the consistency of the base binder.

8.2. Implications for hot climates and transferability

In hot climates such as Iraq and the Middle East, the findings are promising because the High PG, which is used to select the binder in these areas, can be predicted by tests already conducted by local labs. The maximum accuracy with which a designer, knowing only the penetration grade and SP of a refinery binder, can choose a starting SBS content is about one grade (Scenario H-I, Fig. 16). A trial blend tested for Pen and SP reduces the uncertainty

substantially, roughly halving the error from about one grade to half a grade (Figs. 17 and 18; Scenario H-III, LOSO RMSE ≈ 2.9 °C). The leave-one-study-out validation is the most important safeguard of this study. Literature data are hierarchical: records within the same study share the base binder, SBS product, blending equipment, and operator. Random CV rewards models that take these common properties into consideration [77,78]. The fact that the models of Scenario H-III and the three-input nomographs had a LOSO R^2 of 0.75–0.85 indicates that the relationships learned here are not merely a product of the clustering of the records by study, although LOSO does not address selection or publication bias or the dominance of a few large studies. Future experimental work would make their data much more useful for modeling if they reported full Superpave grading with continuous grades, the Pen and SP of the binder prior to and following modification, the SBS product and architecture (with its molecular weight and styrene and vinyl contents), the chemistry of the base bitumen (SARA fractions or colloidal instability index), and the entire blending protocol (temperature, shear rate, and time).

8.3. Limitations

There are several limitations to consider. (i) The database is moderate in size (86 records from 19 studies). The three largest studies account for 41 of the 86 records, and controls make up 41 percent of the data set. (ii) 46 of the 86 High PG values are available only as grades (rounding floor of the RMSE about 1.7 °C), 29 grades were reported by the authors without confirmation of the full M 320 procedure, and 5 records are derived from figures (Section 4.5). (iii) Linear and radial SBS were treated as a single modifier; the paired comparison suggests a possible benefit of radial SBS, but the radial data are too limited to model and generalize. (iv) SBS contents above 8 %, binders with co-modifiers and base binders with softness below 80/100 (Pen_{base} > 100 dmm) were not sampled, and the models are not applicable to them; within the range, the base binder affects the grade both directly (additive terms in Eqs. (10) and (11)) and through the dosage it requires (Fig. 16). (v) The blending protocol (temperature, shear rate, and time), storage stability and phase separation, and possible publication bias were not considered, and the precision of the Pen and SP tests was propagated only into the nomograph readings (Section 7.3), not into the model predictions. (vi) The paper addresses only the M 320 high-temperature grade; the low- and intermediate-temperature requirements and the MSCR-based traffic grade of SBS blends still need to be tested. The database, code, and fitted models are provided so the analysis can be repeated as new data become available.

9. Conclusions

An 86-binder database, including SBS-modified and control binders, was compiled and filtered (19 studies, predominantly from the Middle East). Eight ML algorithms were tested to predict the Superpave High PG using the SBS content and the Pen, SP, and PI of the binder before and/or after modification. Conclusions include:

1. Restricting the database to base binders with penetration grades up to 80/100 yielded 86 binders from 19 studies (PG 58 to PG 100). Within this range, SBS increased the High PG by a median of 2 grades at a median content of 4.5 %, or 3.3–3.9 °C per 1 % SBS on average across the database (about 2.0–2.5 °C per 1 % for a 60/70 binder in the design range). The consistency of the base binder was only weakly correlated with the grade on its own, but substantially once the SBS content was controlled for (partial r = 0.54 for SP_{base}); assuming the grade of the base binder gave negative R^2 in all scenarios.
2. The optimum model (GPR) achieved CV R^2 = 0.85 (RMSE = 3.8 °C, 96 % within one grade) and LOSO R^2 = 0.64 for the SBS content and the consistency of the base binder (H-I). For the consistency of the modified binder (H-II), CV R^2 = 0.70 and 0.68 were obtained with MLR and GPR.
3. Combining consistency between the pre- and post-modification (H-III) produced the most accurate and transferable models: MLR and SVR had CV R^2 = 0.84 and 0.83 (RMSE = 3.0 and 3.1 °C) and LOSO R^2 = 0.85 and 0.72, with 100 % of the CV predictions of MLR within one grade (99.6 % for SVR); for the SBS-modified binders alone, the LOSO RMSE of MLR was 2.7 °C, compared with 5.7 °C for the base grade plus a fitted SBS slope.
4. Permutation importance and SHAP identified the SBS content as the key input before blending and the Pen of the modified binder as the key input after blending; GPR and the boosted trees led when only the base binder was known, linear MLR when the consistency of the blend was known, and the lowest-ranked models were DT in Scenario H-I, ANN in Scenario H-II, and DT in Scenario H-III (Table 12). Grouped out-of-fold importance, which treats Pen, SP, and PI of each binder jointly, confirmed these roles.
5. The GPR design chart indicates mean dosages of about 1.7, 4.4, and 6.1 % SBS to raise a 60/70 binder to PG 70, 76, and 82, and about 3.1 and 5.7 % to raise a 40/50 binder to PG 76 and PG 82; for 80 % reliability, about 2.9 and 5.4 % (60/70) and 2.2 and 5.5 %

(40/50) are needed for PG 70 and PG 76. Dosages above 6.5 % SBS, including PG 88, are not supported by the data.

6. Screening of 89 three-input linear models demonstrated that PI_{base}, log Pen_{mod}, and the SBS content were the best predictors of High PG (R^2 = 0.84, LOSO R^2 = 0.79, 100 % of LOSO predictions within one grade). The model based on the most widely reported test, SP (SP_{base}, SP_{mod}, and SBS content), gave R^2 = 0.80 and LOSO R^2 = 0.75. Both were transformed into alignment nomographs and heat maps that indicate the grade bands across the consistency inputs at 2, 4, and 6 % SBS. The penetration indices prior to and after modification were poor predictors (R^2 = 0.57).
7. Radial and linear SBS can have slightly different High PG values under comparable conditions (2.8 °C on average in 8 paired comparisons), although the small number of radial SBS records does not allow reliable generalization, and the architecture was not entered into the nomographs.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A.H. Albayati, A review of rutting in asphalt concrete pavement, Open Eng. 13 (2023) 20220463, https://doi.org/10.1515/eng-2022-0463.

[2] T.W. Kennedy, G.A. Huber, E.T. Harrigan, R.J. Cominsky, C.S. Hughes, H. Von Quintus, J.S. Moulthrop, Superior Performing Asphalt Pavements (Superpave): The Product of the SHRP Asphalt Research Program, Report SHRP-A-410, Strategic Highway Research Program, National Research Council, Washington, DC, 1994.

[3] G. Zittis, M. Almazroui, P. Alpert, P. Ciais, W. Cramer, Y. Dahdal, et al., Climate change and weather extremes in the Eastern Mediterranean and Middle East, Rev. Geophys. 60 (2022) e2021RG000762, https://doi.org/10.1029/2021RG000762.

[4] W. Zeiada, A.G. Ashour, S.M. Mirou, M. Abuzwidah, et al., Development of superpave asphalt binder specifications to meet climate conditions in the UAE, Int. J. Pavement Eng. 25 (2024) 2386627.

[5] R.I. Al-Nawasir, B.H. Al-Humeidawi, Qualitative evaluation for asphalt binder modified with SBS polymer, Tikrit J. Eng. Sci. 30 (4) (2023) 88–101.

[6] H.M.H. Alani, A.H. Albayati, A.S. Abbas, The transition to a PG grading system for asphalt cement in Iraq, J. Eng. (Baghdad) 16 (4) (2010) 5911–5931, https://doi.org/10.31026/j.eng.2010.04.16.

[7] A.H. Abed, Effects of functionalized polyethylene and styrene butadiene styrene polymers on performance grade of local asphalt binder, J. Eng. (Baghdad) 18 (6) (2012) 735–742.

[8] SCRB, General Specification for Roads and Bridges, Section R/9: Hot-Mix Asphaltic Concrete Pavement, State Corporation of Roads and Bridges, Ministry of Housing and Construction, Baghdad, Iraq, 2003.

[9] X. Yan, D. Wu, K. Hu, W. Zhang, J. Xing, L. Cui, et al., The modification mechanism, evaluation method, and construction technology of direct-to-plant SBS modifiers in asphalt mixture: a review, Polymers 15 (2023) 2768.

[10] M. Bhardwaj, P. Kumar, N. Saboo, Advancements in SBS-modified asphalt: a review of phase structure, modification mechanisms, and storage stability, J. Test. Eval. (2026), https://doi.org/10.1520/JTE20240712.

[11] Y. Li, G. Jiang, B. Xiang, S. Yan, Y. Shen, Z. Ma, et al., Swelling behavior and mechanism of SBS in high viscoelastic-modified bitumen using fluorescence analysis, J. Traffic Transp. Eng. (Engl. Ed.) 13 (4) (2026) 1292–1305.

[12] Y. Kumar, A. Pandey, P. Kumar, S.S. Ravindranath, Elevated temperature rheological properties of styrene-butadiene-modified binders: role of molecular structure, Int. J. Pavement Res. Technol. 16 (2023) 1599–1617.

[13] A. Pandey, S.S. Islam, G.D. Ransinchung R.N., S.S. Ravindranath, Quantifying the effect of SBS molecular structure on the upper service temperature rheological properties of modified binders, Constr. Build. Mater. 352 (2022) 128826.

[14] X. Zhu, Y. Wang, M. Miljković, R. Li, G. Hao, Effects of polymer structure on the physicochemical and performance-related properties of SBS-modified asphalt binders subjected to short-term aging, Constr. Build. Mater. 411 (2024) 134446.

[15] G. Pipintakos, J. Škulteckė, A. Vaitkus, O. Šernas, H. Soenen, W. Van den bergh, Optimising bitumen modification: How styrene–butadiene–styrene (SBS) characteristics and content affect bitumen behaviour, Road Mater. Pavement Des. 26 (sup1) (2025) 443–463.

[16] D. Lesueur, The colloidal structure of bitumen: Consequences on the rheology and on the mechanisms of bitumen modification, Adv. Colloid Interface Sci. 145 (2009) 42–82, https://doi.org/10.1016/j.cis.2008.11.001.

[17] G. Polacco, S. Filippi, F. Merusi, G. Stastna, A review of the fundamentals of polymer-modified asphalts: Asphalt/polymer interactions and principles of compatibility, Adv. Colloid Interface Sci. 224 (2015) 72–112, https://doi.org/10.1016/j.cis.2015.07.010.

[18] W. Wu, M.C. Cavalli, W. Jiang, N. Kringos, Differing perspectives on the use of high-content SBS polymer-modified bitumen, Constr. Build. Mater. 411 (2024) 134433.

[19] AASHTO M 320, Standard Specification for Performance-Graded Asphalt Binder, American Association of State Highway and Transportation Officials, Washington, DC.

[20] AASHTO T 315, Standard Method of Test for Determining the Rheological Properties of Asphalt Binder Using a Dynamic Shear Rheometer (DSR), American Association of State Highway and Transportation Officials, Washington, DC.

[21] AASHTO T 240, Standard Method of Test for Effect of Heat and Air on a Moving Film of Asphalt Binder (Rolling Thin-Film Oven Test), American Association of State Highway and Transportation Officials, Washington, DC.

[22] W. Zeiada, H. Liu, H. Ezzat, G.G. Al-Khateeb, B.S. Underwood, A. Shanableh, et al., Review of the Superpave performance grading system and recent developments in the performance-based test methods for asphalt binder characterization, Constr. Build. Mater. 319 (2022) 126063.

[23] M.A. Ishaq, L. Venturini, F. Giustozzi, Correlation between rheological rutting tests on bitumen and asphalt mix flow number, Int. J. Pavement Res. Technol. 15 (2022) 1297–1316.

[24] AASHTO M 332, Standard Specification for Performance-Graded Asphalt Binder Using Multiple Stress Creep Recovery (MSCR) Test, American Association of State Highway and Transportation Officials, Washington, DC.

[25] AASHTO T 350, Standard Method of Test for Multiple Stress Creep Recovery (MSCR) Test of Asphalt Binder Using a Dynamic Shear Rheometer (DSR), American Association of State Highway and Transportation Officials, Washington, DC.

[26] ASTM D5/D5M, Standard Test Method for Penetration of Bituminous Materials, ASTM International, West Conshohocken, PA.

[27] ASTM D36/D36M, Standard Test Method for Softening Point of Bitumen (Ring-and-Ball Apparatus), ASTM International, West Conshohocken, PA.

[28] M. Girgis, J. Barbier, A. Quignard, I. Merdrignac, J. Marques, Correlation and modelling of the penetration and softening point tests of bitumen binders with a variety of physical and chemical properties, Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles 76 (2021) 57.

[29] C. Wang, L. Song, Z. Wang, Y. Chen, B. Zhou, Correlation analysis of linear viscoelastic (LVE) properties, damage resistance and microstructure of unmodified asphalt binders with the same penetration grade, Materials 15 (2022) 7709.

[30] D. Mirzaiyan, M. Ameri, A. Amini, M. Sabouri, A. Norouzi, Evaluation of the performance and temperature susceptibility of gilsonite- and SBS-modified asphalt binders, Constr. Build. Mater. 207 (2019) 679–692.

[31] R. Justo-Silva, A. Ferreira, G. Flintsch, Review on machine learning techniques for developing pavement performance prediction models, Sustainability 13 (2021) 5248.

[32] N.S.A. Yaro, M.H. Sutanto, M.R. Hainin, N.Z. Habib, A. Usman, M.S. Bello, et al., Soft computing applications in asphalt pavement: a comprehensive review of data-driven techniques using response surface methodology and machine learning, J. Road Eng. 5 (2) (2025) 129–163.

[33] D. Rezazadeh Eidgahee, H. Jahangir, N. Solatifar, P. Fakharian, M. Rezaeemanesh, Data-driven estimation models of asphalt mixtures dynamic modulus using ANN, GP and combinatorial GMDH approaches, Neural Comput. Appl. 34 (2022) 17289–17314.

[34] J. Liu, F. Liu, C. Zheng, D. Zhou, L. Wang, Optimizing asphalt mix design through predicting the rut depth of asphalt pavement using machine learning, Constr. Build. Mater. 356 (2022) 129211.

[35] P. Barghabany, J. Zhang, L.N. Mohammad, S.B. Cooper III, S.B. Cooper Jr., Novel model to predict critical strain energy release rate in semi-circular bend test as fracture parameter for asphalt mixtures using an artificial neural network approach, Transp. Res. Rec. 2676 (2022) 388–400.

[36] Y. Ma, J. Liu, L.N. Mohammad, S.B. Cooper III, S.B. Cooper Jr., M.K.R. Duvvuru, Use of random forest to predict intermediate temperature SCB Jc parameter of long-term aged asphalt mixtures, Transp. Res. Rec. 2678 (2024) 177–189.

[37] A.S. Hosseini, P. Hajikarimi, M. Gandomi, F. Moghadas Nejad, A.H. Gandomi, Optimized machine learning approaches for the prediction of viscoelastic behavior of modified asphalt binders, Constr. Build. Mater. 299 (2021) 124264.

[38] A.M. Al-Sabaeei, H. Alhussian, S.J. Abdulkadir, M. Sutanto, E. Alrashydah, G. Mabrouk, et al., Computational modelling for predicting rheological properties of composite modified asphalt binders, Case Stud. Constr. Mater. 19 (2023) e02651.

[39] R.H. Riyad, R. Jaiswal, I.B. Muhit, J. Shen, Optimizing modified asphalt binder performance at high and intermediate temperatures using experimental and machine learning approaches, Constr. Build. Mater. 449 (2024) 138350.

[40] K. Zhong, Q. Meng, M. Sun, G. Luo, Artificial neural network (ANN) modeling for predicting performance of SBS modified asphalt, Materials 15 (2022) 8695.

[41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[42] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, et al., The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, BMJ 372 (2021) n71, https://doi.org/10.1136/bmj.n71.

[43] Y. Luo, H. Zhang, H. Duan, P. Du, J. Cao, A novel method to improve anti-aging properties of SBS modified bitumen by zinc oxide/expanded vermiculite composite: influence of zinc oxide particle loadings, Constr. Build. Mater. 408 (2023) 133790.

[44] A.H. Nasser, A.H. Abed, Assessment mixing and compaction temperatures for modified HMA using Superpave high shear viscosity methods, Al-Nahrain J. Eng. Sci. 21 (4) (2018) 516–522.

[45] M.D.I. Domingos, A.L. Faxina, Accelerated short-term ageing effects on the rheological properties of modified bitumens with similar high PG grades, Road Mater. Pavement Des. 16 (2) (2015), https://doi.org/10.1080/14680629.2014.995209.

[46] M.D.I. Domingos, A.L. Faxina, Avaliação do comportamento fluência-recuperação de ligantes asfálticos modificados com SBS e ácido polifosfórico, Transportes 21 (1) (2013) 51–58.

[47] F. Ozel, M.T. Deniz, M.I. Yuce, Evaluation of olive pomace and SBS modified bitumen to the performance characteristics, Case Stud. Constr. Mater. 19 (2023) e02432.

[48] A. Sağlık, A.G. Güngör, F. Orhan, O. Özay, E. Arıkan Öztürk, Evaluation of performance grades and polymer dispersion of polymer modified binders, in: Proc. 5th Eurasphalt & Eurobitume Congress, Istanbul, 2012, Paper A5EE-457.

[49] C. Dong, L. Feng, Y. Xu, Performance zoning of asphalt pavement and performance grade (PG) of asphalt binder in Karamay: a case study of Xinjiang, China, Sustainability 15 (2023) 9742.

[50] F.M. Jasim, Merits of polymer types used with different local bitumen produced in Kurdistan-Iraq, Research paper, Erbil Polytechnic University, Erbil, Iraq, 2016.

[51] A. Alnaqbi, G. Al-Khateeb, W. Zeiada, et al., Effect of styrene-butadiene-styrene on the properties and grading of local asphalt binder in the UAE, IOP Conf. Ser. Mater. Sci. Eng. 1044 (2021) 012003.

[52] E. Çavdar, A. Kumandaş, N. Şahan, E. Kabadayı, Ş. Oruç, Investigation of rheological and physical properties of SBS and WCO composite modified bitumen, Turk. J. Civ. Eng. (2024), Paper 785, 1–18.

[53] A.O. Mahmood, R.A. Kattan, Improvement of engineering properties of asphalt binder and mixture by using SBS additive material, Adv. Civ. Eng. 2023 (2023) 7151175.

[54] H.A. Alsolieman, A.M. Babalghaith, Z.A. Memon, A.S. Al-Suhaibani, A. Milad, Evaluation and comparison of mechanical properties of polymer-modified asphalt mixtures, Polymers 13 (2021) 2282.

[55] A.H. Albayati, M.J. Al-Kheetan, A.F. Al-Ani, Y. Wang, A.M. Mohammed, M.M. Moudhafar, Effects of polyphosphoric acid on physical, rheological, and chemical properties of styrene-butadiene-styrene (SBS)-modified asphalt binder, J. Compos. Sci. 9 (2) (2025) 78.

[56] C. Yan, W. Huang, J. Ma, J. Xu, Q. Lv, P. Lin, Characterizing the SBS polymer degradation within high content polymer modified asphalt using ATR-FTIR, Constr. Build. Mater. 233 (2020) 117708.

[57] A. Eltwati, Z. Al-Saffar, A. Mohamed, M.R. Hainin, A. Elnihum, M. Enieb, Synergistic effect of SBS copolymers and aromatic oil on the characteristics of asphalt binders and mixtures containing reclaimed asphalt pavement, Constr. Build. Mater. 327 (2022) 127026.

[58] J.Ph. Pfeiffer, P.M. Van Doormaal, The rheological properties of asphaltic bitumens, J. Inst. Petrol. 22 (1936) 414–440.

[59] J. Read, D. Whiteoak, The Shell Bitumen Handbook, 5th ed., Thomas Telford, London, 2003.

[60] ASTM D7643, Standard Practice for Determining the Continuous Grading Temperatures and Continuous Grades for PG Graded Asphalt Binders, ASTM International, West Conshohocken, PA.

[61] F. Zhang, A. Cannone Falchetto, D. Wang, Z. Li, Y. Sun, W. Lin, Prediction of asphalt rheological properties for paving and maintenance assistance using explainable machine learning, Fuel 396 (2025) 135319.

[62] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (2004) 199–222, https://doi.org/10.1023/B:STCO.0000035301.49549.88.

[63] X. Fan, S. Lv, C. Xia, D. Ge, C. Liu, W. Lu, Strength prediction of asphalt mixture under interactive conditions based on BPNN and SVM, Case Stud. Constr. Mater. 21 (2024) e03489.

[64] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, Cambridge, MA, 2006.

[65] M. Khadijeh, C. Kasbergen, S. Erkens, A. Varveri, Combining deep neural networks and Gaussian processes for asphalt rheological insights, Results Eng. 26 (2025) 105629.

[66] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[67] H. Khani Sanij, R. Babagoli, R. Mohammadi Elyasi, Enhancing stone matrix asphalt performance with sugarcane bagasse ash: Mechanical properties and machine learning-based predictions using XGBoost and random forest, Case Stud. Constr. Mater. 23 (2025) e05186.

[68] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in: Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.

[69] A. Gupta, S. Gowda, A. Tiwari, A.K. Gupta, XGBoost-SHAP framework for asphalt pavement condition evaluation, Constr. Build. Mater. 426 (2024) 136182.

[70] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, LightGBM: a highly efficient gradient boosting decision tree, Adv. Neural Inf. Process. Syst. 30 (2017) 3146–3154.

[71] H. Ha, T.T.T. Trang, D.D. Nguyen, A. Islam, H.T. Tran, I. Prakash, et al., Hybrid model of LightGBM regression and grid search optimization for the estimation of permanent deformation of asphalt mixtures pavements, Eng. Technol. Appl. Sci. Res. 15 (2025) 22901–22907.

[72] E.S. Muckley, J.E. Saal, B. Meredig, C.S. Roper, J.H. Martin, Interpretable models for extrapolation in scientific machine learning, Digit. Discov. 2 (2023) 1425–1435.

[73] B. Bischl, M. Binder, M. Lang, T. Pielok, J. Richter, S. Coors, et al., Hyperparameter optimization: foundations, algorithms, best practices, and open challenges, WIREs Data Min. Knowl. Discov. 13 (2023) e1484.

[74] N. Baldo, F. Rondinella, F. Daneluz, P. Vacková, J. Valentin, M.D. Gajewski, et al., Stiffness moduli modelling and prediction in four-point bending of asphalt mixtures: A machine learning-based framework, CivilEng 4 (2023) 1083–1097.

[75] S. Bates, T. Hastie, R. Tibshirani, Cross-validation: what does it estimate and how well does it do it?, J. Am. Stat. Assoc. 119 (2024) 1434–1445.

[76] J. Wainer, G. Cawley, Nested cross-validation when selecting classifiers is overzealous for most practical applications, Expert Syst. Appl. 182 (2021) 115222.

[77] S. Kapoor, A. Narayanan, Leakage and the reproducibility crisis in machine-learning-based science, Patterns 4 (2023) 100804, https://doi.org/10.1016/j.patter.2023.100804.

[78] S. Durdy, M.W. Gaultois, V.V. Gusev, D. Bollegala, M.J. Rosseinsky, Random projections and kernelised leave one cluster out cross validation: universal baselines and evaluation tools for supervised machine learning of material properties, Digit. Discov. 1 (2022) 763–778.

[79] T.-T.-H. Hoang, H.-L. Nguyen, H.-G.T. Hoang, N.K. Bui, H.-B. Ly, Interpretable machine learning framework for predicting air void content in sustainable steel-slag SMA mixtures using metaheuristic-optimized XGBoost, Buildings 16 (2026) 3564.

[80] O. Rainio, J. Teuho, R. Klén, Evaluation metrics and statistical tests for machine learning, Sci. Rep. 14 (2024) 6086.

[81] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Stat. 6 (1979) 65–70.

[82] K.E. Taylor, Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res. 106 (D7) (2001) 7183–7192, https://doi.org/10.1029/2000JD900719.

[83] A. Fisher, C. Rudin, F. Dominici, All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously, J. Mach. Learn. Res. 20 (177) (2019) 1–81.

[84] G. Hooker, L. Mentch, S. Zhou, Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance, Stat. Comput. 31 (2021) 82.

[85] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017) 4765–4774.

[86] D.S. Davis, Nomography and Empirical Equations, Reinhold Publishing, New York, 1955.

[87] A.C. Cameron, D.L. Miller, A practitioner's guide to cluster-robust inference, J. Hum. Resour. 50 (2015) 317–372, https://doi.org/10.3368/jhr.50.2.317.

[88] AASHTO M 323, Standard Specification for Superpave Volumetric Mix Design, American Association of State Highway and Transportation Officials, Washington, DC.

[89] S.K. Singh, Y. Kumar, S.S. Ravindranath, Thermal degradation of SBS in bitumen during storage: influence of temperature, SBS concentration, polymer type and base bitumen, Polym. Degrad. Stab. 147 (2018) 64–75.

[90] G.D. Airey, Rheological properties of styrene butadiene styrene polymer modified road bitumens, Fuel 82 (2003) 1709–1719, https://doi.org/10.1016/S0016-2361(03)00146-7.

[91] X. Lu, U. Isacsson, Rheological characterization of styrene-butadiene-styrene copolymer modified bitumens, Constr. Build. Mater. 11 (1997) 23–32.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SBSPGDatabase.xlsx
- SBSPGmodelparameters.xlsx
- SBSPGresultstables.xlsx
- figS3predHIII.png
- figS2predHII.png
- figS1predHI.png