Journal of Industrial and Management Optimization

https://www.aimspress.com/journal/jimo

JIMO, 22(1): 256–281.
DOI:10.3934/jimo.2026010
Received: 30 July 2025
Revised: 19 September 2025
Accepted: 09 October 2025
Published: 21 November 2025

Research article

A multi-objective framework based on estimation of distribution algorithms for data-driven fuel experimental design

Vicente P. Soloviev[1,*], Pedro Larrañaga[1], Marco Bernabei[2], Marina A. Chirita[2], Jose M. Seoane[2], Pedro Fontán[2], and Concha Bielza[1]

[1] Universidad Politécnica de Madrid, Madrid, Spain

[2] Repsol Technology Lab., Madrid, Spain

* **Correspondence:** Email: vicente.perez.soloviev@upm.es.

Supplementary

Ingredients and properties descriptions

This section describes the initial historical data provided by the experts, used for the EDA approach initialization.

Table 1 shows the properties $P1, \ldots, P14$ associated to each of the $n = 24$ ingredients for the fuel fabrication. Note that the names of the ingredients are denoted referring to the category to which they belong. For example, $A1$ is the first ingredient of category A and $B2$ the second one of category B.

Figure 1 shows a histogram for each of the properties $P1, \ldots, P14$.

Table 1. Properties $P1, \ldots, P14$ associated to each of the ingredients $I_1, \ldots, I_{24}$ used for the fuel fabrication and its name denoting the category to which they belong.

		P1	*P2*	*P3*	*P4*	*P5*	*P6*	*P7*	*P8*	*P9*	*P10*	*P11*	*P12*	*P13*	*P14*
I_1	*A1*	678.6	76.5	42.0	100.0	100.0	0.0	0.0	99.4	0.0	85.6	14.4	6.0	12.0	0.0
I_2	*A2*	667.0	97.7	102.2	100.0	100.0	0.0	0.0	99.9	0.1	85.6	14.4	5.0	10.0	0.0
I_3	*A3*	816.3	83.9	26.8	0.0	100.0	0.0	0.0	99.9	0.1	87.7	12.3	6.0	10.0	0.0
I_4	*A4*	777.1	93.4	68.4	100.0	100.0	0.0	0.0	98.9	0.0	87.6	11.8	4.9	7.9	0.0
I_5	*B1*	652.6	85.6	78.5	100.0	100.0	0.0	0.0	0.0	0.0	68.7	13.5	4.3	10.1	0.0
I_6	*B2*	652.5	88.6	78.4	100.0	100.0	0.0	0.0	0.0	0.0	83.6	16.4	5.7	13.3	0.0
I_7	*C1*	784.1	83.0	22.0	0.0	100.0	0.0	0.0	0.0	0.0	85.6	14.4	6.0	12.0	0.0
I_8	*C2*	746.6	100.6	66.0	100.0	100.0	0.0	0.0	0.0	0.0	85.5	14.5	5.0	10.1	0.0
I_9	*C3*	754.0	91.0	31.6	0.0	100.0	0.0	0.0	0.0	0.1	85.6	14.4	6.0	12.0	0.0
I_{10}	*D1*	723.0	105.0	33.0	100.0	100.0	0.0	15.7	0.0	0.0	70.5	13.8	6.0	14.0	1.0
I_{11}	*D2*	751.6	113.4	35.7	32.0	100.0	100.0	14.7	3.3	0.1	71.1	13.9	5.8	13.6	0.9
I_{12}	*D3*	751.6	113.4	35.7	32.0	100.0	50.0	14.7	3.3	0.1	71.1	13.9	5.8	13.6	0.9
I_{13}	*D4*	793.7	105.4	17.1	0.0	100.0	100.0	34.7	0.0	0.0	52.1	13.1	2.0	6.0	1.0
I_{14}	*D5*	800.0	105.1	8.0	0.0	0.0	0.0	21.6	0.0	0.0	64.8	13.6	4.0	10.0	1.0
I_{15}	*D6*	746.6	117.0	60.1	100.0	100.0	39.0	18.1	0.0	0.0	68.1	13.7	5.0	12.0	1.0
I_{16}	*D7*	796.9	106.0	32.8	100.0	100.0	100.0	49.9	0.0	0.0	37.5	12.6	1.0	4.0	1.0
I_{17}	*D8*	770.9	114.0	19.5	0.0	100.0	0.0	15.1	0.8	0.3	70.9	13.8	5.8	13.4	0.9
I_{18}	*E1*	625.2	91.5	135.1	100.0	100.0	0.0	0.0	0.0	0.0	83.2	16.8	5.0	12.0	0.0
I_{19}	*E2*	699.4	96.7	12.7	0.0	64.0	0.0	0.0	0.0	0.0	81.9	15.5	7.7	17.4	0.0
I_{20}	*E3*	707.3	77.4	30.3	26.3	67.4	100.0	0.0	2.5	1.9	50.8	9.0	3.1	6.5	0.0
I_{21}	*E4*	756.8	90.3	54.4	23.9	40.0	100.0	0.0	5.3	38.1	89.7	10.3	3.3	6.9	0.0
I_{22}	*F1*	871.4	107.4	3.4	0.0	0.0	0.0	0.0	0.0	100.0	90.5	9.5	8.0	10.0	0.0
I_{23}	*F2*	871.4	120.1	6.4	0.0	0.0	0.0	0.0	0.0	99.9	91.1	8.7	7.0	8.0	0.0
I_{24}	*F3*	868.4	117.5	2.7	0.0	0.0	0.0	0.0	0.0	100.0	90.5	9.5	8.0	10.0	0.0

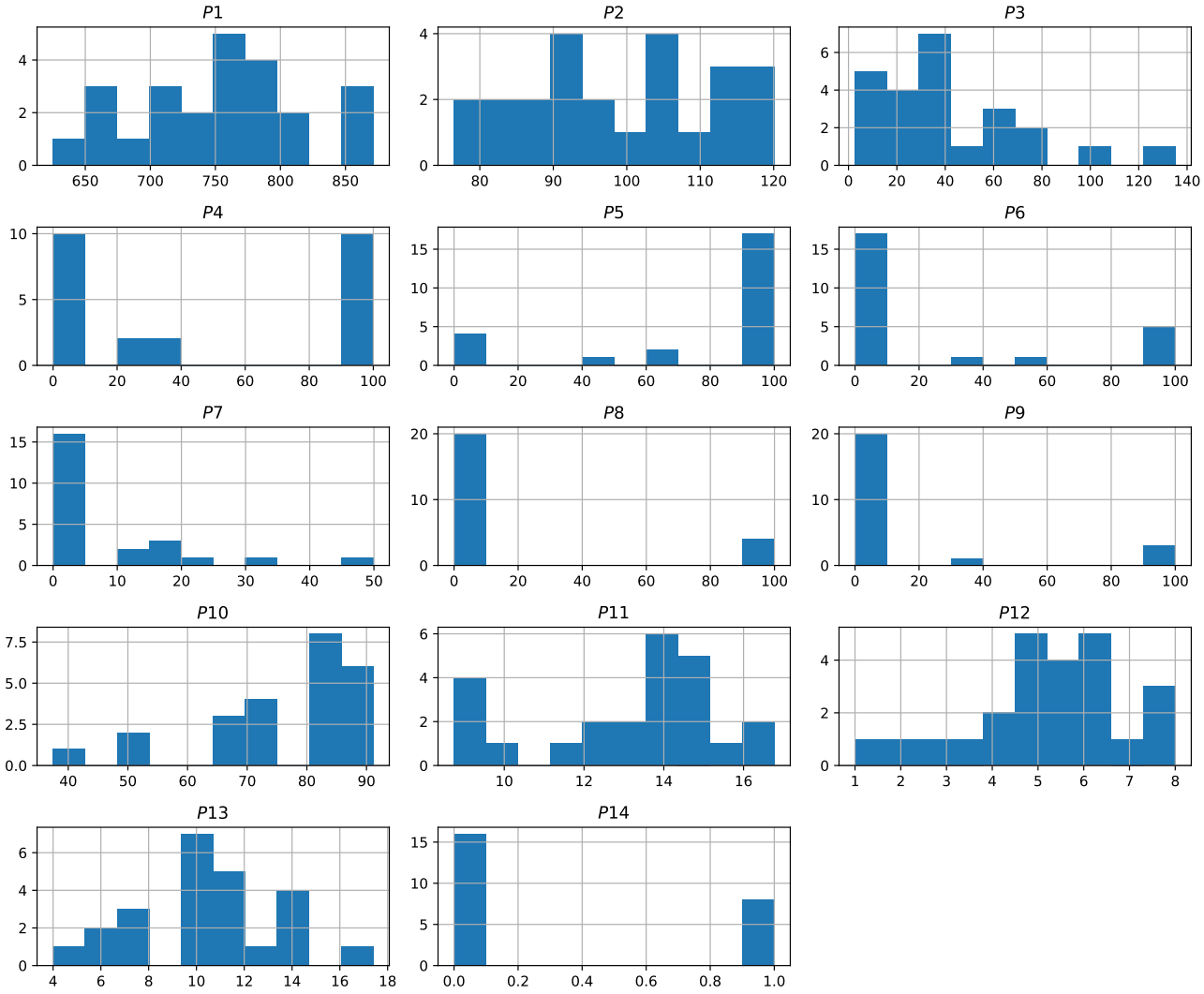

Figure 1. Frequency of each property $P1, \ldots, P14$.

Optimization constraints

Table 2 shows the lower and upper bounds, respectively, for analytical functions $W^i_{calc}(i = 1, \ldots, 7)$ and laboratory experimentation $W^i_{lab}(i = 1, \ldots, 4)$.

Configuration of regression models

In this section, we detail the hyperparameter configurations applied to the regression models tested to predict W^1_{lab}. Tables 3–4 show the hyperparameters used in the experimental results for Bayesian ridge, LASSO, ridge and kernel ridge models. Note that categorical hyperparameters were used as the default in scikit-learn-1.3.0 [52]. The Bayesian ridge model was selected as the best regression model to predict W^1_{lab} values, so both the default and after-tuning hyperparameters are shown. Hyperparameter tuning was performed by Bayesian optimization. All the experiments use the same maximum number of iterations set as the default.

Table 2. Formula, lower bound (LB), and upper bound (UB) associated to each analytical descriptor W^i_{calc} and each laboratory descriptor W^i_{lab}; where x_i is the percentage of each ingredient I_i in the total mixture $\boldsymbol{x}$, and $Pj(I_i)$ is the property j of ingredient I_i (see Table 1).

	Formula	LB	UB
$W^1_{calc}(\boldsymbol{x})$	$\frac{100}{\sum_{i=1}^{n}(x_i/P1(I_i))}$	LB^1_{calc}	UB^1_{calc}
$W^2_{calc}(\boldsymbol{x})$	$\sum_{i=1}^{n}\left(x_i \frac{1000(\frac{32.8P10(I_i)}{100}+141.8(\frac{P11(I_i)}{100}-\frac{P7(I_i)}{800}))-\frac{2440*9P11(I_i)}{100}}{\frac{P12(I_i)+P13(I_i)/4-P14(I_i)/2}{6.09/(12P12(I_i)+P13(I_i)+16P14(I_i))}}\right)$	LB^2_{calc}	UB^2_{calc}
$W^3_{calc}(\boldsymbol{x})$	$\frac{\sum_{i=1}^{n} x_i P8(I_i)/P1(I_i)}{\sum_{i=1}^{n} x_i/P1(X_i)}$	LB^3_{calc}	UB^3_{calc}
$W^4_{calc}(\boldsymbol{x})$	$\sum_{i=1}^{n} x_i P7(I_i)$	LB^4_{calc}	UB^4_{calc}
$W^5_{calc}(\boldsymbol{x})$	$\frac{\sum_{i=1}^{n} x_i P9(I_i)/P1(I_i)}{\sum_{i=1}^{n} x_i/P1(I_i)}$	LB^5_{calc}	UB^5_{calc}
$W^6_{calc}(\boldsymbol{x})$	$\sum_{i=1}^{n} x_i P6(I_i)$	LB^6_{calc}	UB^6_{calc}
$W^7_{calc}(\boldsymbol{x})$	x_{16}	LB^7_{calc}	UB^7_{calc}
$W^1_{lab}(\boldsymbol{x})$	-	LB^1_{lab}	UB^1_{lab}
$W^2_{lab}(\boldsymbol{x})$	-	LB^2_{lab}	UB^2_{lab}
$W^3_{lab}(\boldsymbol{x})$	-	LB^3_{lab}	UB^3_{lab}
$W^4_{lab}(\boldsymbol{x})$	-	LB^4_{lab}	UB^4_{lab}

Table 3. Bayesian ridge model hyperparameters by default and after hyperparameter tuning (hyperopt), where *tol* is the tolerance error to consider convergence; α_1 and α_2 are the shape parameter for the Gamma distribution and the inverse scale parameter for α, respectively; and λ_1 and λ_2 are the shape parameter for the Gamma distribution and the inverse scale parameter for λ, respectively.

	tol	α_1	α_2	λ_1	λ_2
default	1e-3	1e-6	1e-6	1e-6	1e-6
hyperopt	94.64	50.16	18.59	81.89	0.20

Table 4. The LASSO, ridge and kernel ridge regression models' hyperparameter configurations used. Parameter α_3 is the constant that multiplies the L1 and L2 terms in LASS and ridge regression models, respectively, and *tol* represents the tolerance error to consider algorithm convergence. In the case of the kernel ridge, *degree* is the polynomial degree used.

	α_3	*tol*	degree
Lasso	0.5	1e-4	-
Ridge	0.5	1e-4	-
Kernel ridge	10	-	3

Competitors' configurations

In this section, we detail the hyperparameter configuration applied to the other optimizers used in the comparison against the EDA approach presented in this article. NSGA-II and MOEAD implementations have been obtained from the PYMOO Python library and both algorithms were initialized from the historical data. Tables 5–6 show the hyperparameter configurations used for both algorithms in the experimental results.

Table 5. Hyperparameters used for the NSGAII algorithm [43], where *N* is the population size; *offs* regards the number of *offsprings* performed; *P* is the probability of crossover and mutation occurrence; and the *eta* parameter controls the shape of the distribution used for both crossover and mutation operators.

			Crossover		Mutation	
	N	*offs*	*P*	*eta*	*P*	*eta*
tested	200	10	0.9	15	1.0	20

Table 6. Hyperparameters used for MOEAD algorithm [56], where *n_neig* is the number of neighbors considered and *P* is the probability of neighbor mating occurrence.

	n_neig	*P*
tested	15	0.7

AIMS Press