



HAL
open science

A digital twin integrating multi-objective optimization to support fryer operators in managing potato crisps production

Manon Perrignon, Mathieu Emily, Mélanie Munch, Paul Debuire, Romain Jeantet, Thomas Croguennec

► **To cite this version:**

Manon Perrignon, Mathieu Emily, Mélanie Munch, Paul Debuire, Romain Jeantet, et al.. A digital twin integrating multi-objective optimization to support fryer operators in managing potato crisps production. *Journal of Food Engineering*, 2026, 406, pp.112800. <10.1016/j.jfoodeng.2025.112800>. <hal-05247019>

HAL Id: hal-05247019

<https://hal.science/hal-05247019v1>

Submitted on 9 Sep 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Journal Pre-proof

A digital twin integrating multi-objective optimization to support fryer operators in managing potato crisps production

Manon Perrignon, Mathieu Emily, Mélanie Munch, Paul Debuire, Romain Jeantet, Thomas Croguennec



PII: S0260-8774(25)00335-8

DOI: <https://doi.org/10.1016/j.jfoodeng.2025.112800>

Reference: JFOE 112800

To appear in: *Journal of Food Engineering*

Received Date: 17 March 2025

Revised Date: 29 August 2025

Accepted Date: 3 September 2025

Please cite this article as: Perrignon, M., Emily, M., Munch, M., Debuire, P., Jeantet, R., Croguennec, T., A digital twin integrating multi-objective optimization to support fryer operators in managing potato crisps production, *Journal of Food Engineering*, <https://doi.org/10.1016/j.jfoodeng.2025.112800>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2025 Published by Elsevier Ltd.

1 **Title:** A digital twin integrating multi-objective optimization to support fryer operators in managing
2 potato crisps production

3 **Authors:** Manon Perrignon¹, Mathieu Emily², Mélanie Munch³, Paul Debuire⁴, Romain Jeantet¹ and
4 Thomas Croguennec^{1*}

5 ¹ L'Institut Agro, INRAE, STLO (Science et Technologie du Lait et de l'œuf), Rennes, France

6 ² L'Institut Agro, Université de Rennes, CNRS, IRMAR (Institut de Recherche Mathématique de
7 Rennes)-UMR 6625, Rennes, France

8 ³ INRAE, STLO (Science et Technologie du Lait et de l'œuf), Rennes, France

9 ⁴ ALTHO - Brets, Saint Gérard, France

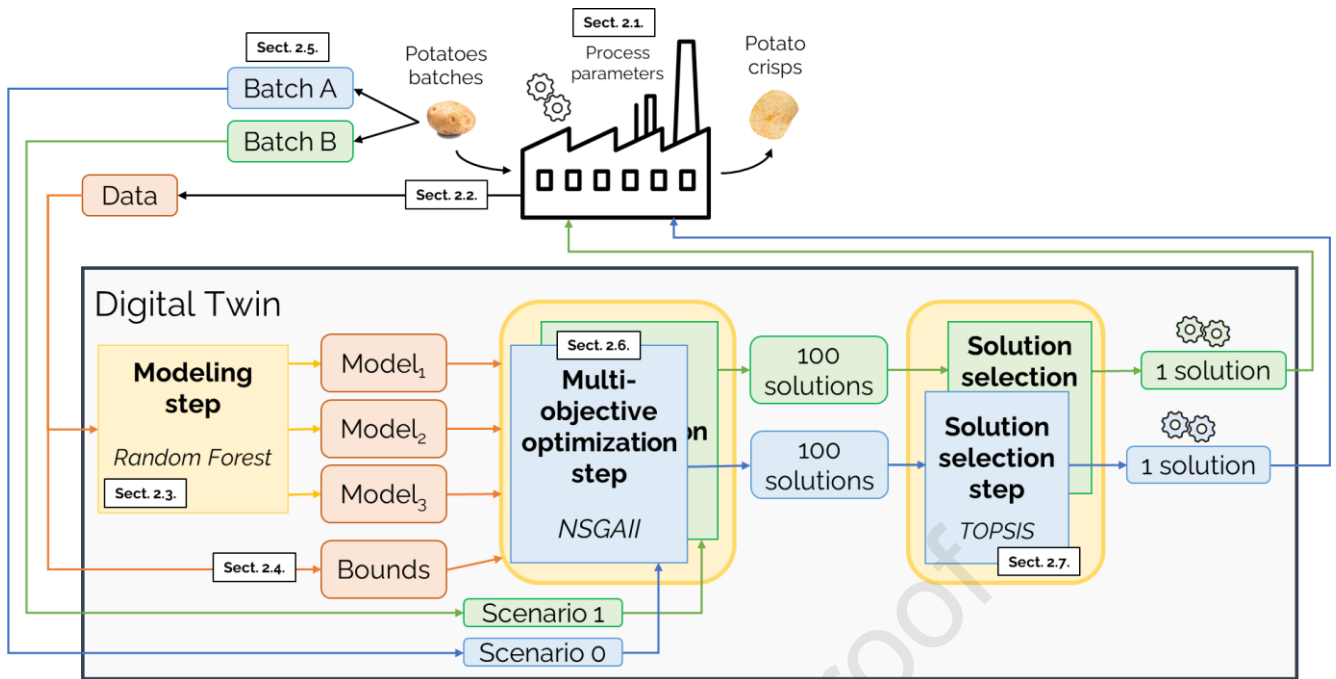
10 *Corresponding author

11 **Keywords:** Digital Twin, Multi-objective optimization, Machine Learning, Potato crisps, Decision-
12 support tool

13 **Abstract:**

14 The production of potato crisps currently relies on the expertise of human operators, known as fryers,
15 whose training is long and demanding. With the decline in fryer vocations and the increase in
16 consumer quality expectations, it is now essential to develop decision-support tools to make fryer
17 work easier and better control product quality. This study proposes a digital twin (DT) approach that
18 incorporates multi-objective optimization to assist fryers in managing their crisp production line. Data
19 from crisp production is collected and used to model key crisp physicochemical indicators (Fat
20 content, Moisture content, and Lightness) using a Machine Learning approach, specifically the
21 Random Forest method. Then, a multi-objective optimization is carried out using the Non-dominated
22 Sorting Genetic Algorithm II (NSGA-II) algorithm, which identifies optimal adjustments to enhance
23 the three crisp physicochemical indicators. The optimization is tested on 2 batches of potatoes as
24 inputs. The algorithm generates a set of optimal solutions, from which a final solution is selected
25 using the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) multi-criteria
26 decision-making method. This solution provides practical recommendations for adjusting production
27 parameters according to a given potato batch quality. The results show that physicochemical
28 parameters of the crisps are similar after optimization, regardless of the quality the potato batch.
29 Variation in potato batch quality is compensated by appropriate adjustments of the crisp
30 manufacturing process parameters ensuring consistent and optimal production. In conclusion, this
31 digital twin, integrating multi-objective optimization, proves to be a valuable tool for improving fryer
32 decision-making and optimizing production line management.

33 **Graphical abstract:**



34

35 1 Introduction

36 The manufacturing process of potato crisps is a complex process, which is monitored by skilled
 37 operators called fryers, who adjust the various parameters of the production line. The fryers have
 38 extensive technical expertise they have developed through rigorous and specialized training.
 39 However, the dwindling number of qualified candidates due to the demanding nature of the training
 40 underlines the urgent need to offer new support tools to the fryers. The fryers must control the process
 41 to meet several crisps quality criteria in terms of texture, visual appearance, and taste (Salvador et al.,
 42 2009). These objectives are linked to specific physicochemical indicators of potato crisp, among
 43 which fat and moisture contents of the crisps, and crisp lightness (Zhang et al., 2024). Fat content
 44 contributes to both flavor and texture. Moisture content influences crisps shelf life and texture,
 45 particularly its crispiness (Glasse et al., 2017). If the moisture content is too high or too low, it can
 46 lead to either soggy or overly dry crisps, respectively. Crisp lightness plays a critical role in the visual
 47 appeal of the crisps. Deviations in crisp lightness can lead to non-conformity, particularly if the crisps
 48 appear excessively brown. This browning is primarily due to Maillard reaction, which also generates
 49 small amount of acrylamide, a carcinogen compound formed during the high-temperature cooking of
 50 starchy foods and aroma compounds that affects the sensory properties of the crisps (Martinez et al.,
 51 2019). The lightness is evaluated as a percentage of lightness (L^*). In the CIELAB color space,
 52 lightness quantifies the visual perception of color brightness on a scale from 0 (black) to 100 (white),
 53 where higher values correspond to lighter surface coloration. A higher value is correlated with lower
 54 acrylamide content (Viklund et al., 2007) and more appealing crisps.

55 These indicators represent key measures of the industry's performance, which fryers must optimize
56 while adapting to the variability of potatoes (Peñalver-Soto et al., 2022). However, optimizing this
57 performance is particularly challenging, as it involves balancing multiple conflicting objectives
58 (Drofenik et al., 2023), such as crisp lightness and moisture content. To address this complexity,
59 multi-objective optimization (Gunantara, 2018) (Wari and Zhu, 2016) helps solve such food process
60 optimization problem by simultaneously dealing with multiple performance objectives, identifying
61 trade-offs, and providing the ideal parameters to reach the objectives set (Konak et al., 2006). The
62 multi-objective optimization of processes is a well-defined procedure, but the implementation of the
63 different steps depends on the specific problem to be solved and the application domain (Perrignon
64 et al., 2024). For instance, the step of modeling objectives can be approached in various ways. One
65 option is the use of available mathematical equations, known as mechanistic modeling, while another
66 is data-driven model (Madoumier et al., 2019). There are no global equations to describe the entire
67 process of potato crisp manufacture. The process is controlled by setting/measuring various
68 manufacturing parameters but the relationships between these parameters are often complex
69 justifying the use of data-driven methods, such as Machine Learning. These methods allow for
70 establishing relationships between collected manufacturing parameters to predict the defined
71 objectives (Garre et al., 2020). However, their effectiveness relies on the availability of sufficient and
72 relevant data. Currently, a large amount of data is collected throughout the potato crisp production
73 process, providing an opportunity to apply these methods. It has been shown that data collected from
74 complex food transformation processes can be leveraged to better manage food processing
75 performance indicators and gain a deeper understanding of the influencing process parameters
76 (Perrignon et al., 2025).

77 These data can also be integrated in virtual replicas of food processes providing innovative solutions
78 for managing food processes in real-time, while optimizing performance and product quality
79 (Henrichs et al., 2022). These so-called digital twins (DT) (Shankarrao Patange and Bharatkumar
80 Pandya, 2022) are running in parallel to the real food manufacturing process through a continuously
81 updated data flow (Verboven et al., 2020). By integrating data from sensors installed all along the
82 food process, DT enable continuous performance monitoring and facilitate rapid, data-driven
83 decision-making. When a change is measured in the food manufacturing process, DT integrates the
84 information and sends updated instructions to the real food process (Kritzinger et al., 2018). This
85 technology can be leveraged to optimize production lines, enhance product quality and reduce costs
86 (Barricelli et al., 2019). It has therefore the potential to improve the management of crisps
87 manufacturing process, assisting fryers in making decisions towards constant crisps quality
88 achievement, simplifying their training and making their role more accessible.

89 The aim of this paper is to propose a Digital Twin approach incorporating a multi-objective
90 optimization and data-driven method to provide a decision support tool to the fryers. To achieve this,
91 the data collected through the whole process from the potato batches to the crisps before salting and
92 packaging were used to model crisp physicochemical indicators that were then used for multi-
93 objective optimization. To illustrate the optimization approach two batches of potato of different
94 quality (A and B) were used as inputs. The optimization returns the manufacturing parameters to be
95 implemented by the fryers for the two batches of potatoes to meet the crisps quality objectives. This
96 study showed that, depending on the quality of the potatoes, the process has to be managed differently,
97 and implementing this approach through a Digital Twin could assist and guide fryers in setting the
98 process parameters for optimizing crisp quality.

99 **2 Methodology**

100 **2.1 Manufacturing process overview**

101 The potato crisp manufacturing process under study is illustrated in Figure 1. The potatoes in batches
102 are stored in hoppers. A batch represent a defined volume of potatoes. For each batch, the dry matter
103 content and the size of the potatoes are measured. The potatoes are classified as very small, small,
104 medium or big. The potatoes are then sent to an optical sorter to identify any appearance defects:
105 bruises, deformations, percentage of burnt, rotten, spotted and green, sprouts on the potatoes. They
106 are washed before being sent to a Pulsed Electric Field (PEF) system, which helps reducing the final
107 acrylamide content in crisps (Bruno et al., 2024). The potatoes are then peeled, calibrated, and
108 possibly cut to achieve a uniform size. A buffer zone is added to regulate the potato flow before they
109 pass through an elevator to be sliced. The slats of potatoes are washed, drained, and then sent to the
110 fryer. The fryer consists of three belts circulating the potato slats through the frying oil. After exiting
111 the fryer, the crisps go through an optical sorter and an infrared spectrometer to remove non-conform
112 crisps and measure the indicators that define the crisps (fat and moisture contents, lightness). The
113 crisps are then sent to the salt shaker before being placed in a buffer zone before packaging. Each red
114 star in Figure 1 represents a data collection point.

115 **2.2 Data collection and preprocessing**

116 The data come from two years of production (2022-2023) of one production line of a crisp
117 manufacturing industry. Throughout the production process, from potato storage to crisp packaging,
118 various manufacturing parameters are recorded using automated sensors (red stars in Figure 1) and
119 collected in a data file. These parameters include those defining the raw material (quality criteria of

120 the potato batches), production conditions (peeling time, frying temperature, crisp thickness, etc.), as
121 well as physicochemical indicators of the potato crisps. The data file is composed of 135,849
122 individuals (potato batches) and 47 variables (manufacturing parameters). A simplified form of the
123 dataset is shown in Figure 2 and a detailed summary of the variables is provided as supplementary
124 material (Supplementary material Table S2).

125 A batch of potatoes, which is defined by quality variables, is processed under uniform conditions.
126 This means that in the dataset, each batch of potatoes is defined by 47 variables, specifically
127 comprising 19 potato quality variables, 25 process parameters and 3 crisp physicochemical indicators
128 (fat and moisture content, and lightness). This dataset links the quality of the potatoes, the setting of
129 the manufacturing process parameters and the physicochemical indicators of the potato crisps.

130 Variables have different roles in the process and can be classified into three categories. The first
131 category gathers the 3 performance indicators (fat content (%), moisture content (%) and lightness
132 (%)) that constitute the crisp quality objectives. The second category gathers 9 decision variables that
133 refer to actionable parameters in the process, meaning they can be directly adjusted by the fryers to
134 influence the course and performance of the process. These variables represent the levers on which
135 the DT will provide decision-making support. The identified variables, listed in chronological order
136 along the crisp process, are: Peeling Time, Elevator drive sprocket frequency, Slat Thickness, Water
137 Supply, Boiler Temperature, Fryer Temperature, Time Belt 1, Time Belt 2 and Belt 3 drive sprocket
138 frequency. The third category includes 35 informative variables, which are non-actionable parameters
139 defining the process trajectory (e.g., potato quality variables, other process variables) (Figure 2).

140 A data preprocessing phase is undertaken to transform raw data into an exploitable database. This
141 phase allows for data cleaning by removing missing values, duplicates, or errors, standardizing
142 formats and measurement units, and structuring information consistently. Without this preparation,
143 results may be biased or unreliable, since their quality directly depends on the quality of the input
144 data. This data preprocessing phase is divided into three steps.

145 The first step consists of detecting outliers. To achieve this, data visualization techniques such as box
146 plots or distribution graphs were used. This filtering is performed using expert knowledge based on
147 acceptance thresholds (Supplementary material, Figure S1). These thresholds were established for
148 each parameter, defining the standard operating range of the manufacturing process. Experts are again
149 consulted to determine whether the detected outliers provide useful information or are simply data
150 entry errors. Their knowledge allows for evaluating the nature of these outliers. Data collection
151 problems (automatic data recording when manufacturing is stopped) were considered as outliers.
152 Individuals in the dataset presenting outliers were removed. The second step consists of removing
153 redundant variables, as some variables may be represented with different units (for example, minutes

154 or seconds), or variables that do not fall within the scope of the study (for example, final product
155 traceability). The final step consists of transforming the initial dataset. For this study, we chose to
156 normalize the data by centering and scaling the data so that all variables have the same weight.

157 2.3 Definition and modeling of the performance indicator with Random Forest

158 The three crisp performance indicators (fat and moisture contents and lightness) were modeled using
159 several methods of machine learning (details of methods are provided in supplementary material).
160 For each method and each performance indicator, the complete dataset was divided into two parts:
161 80% of the data was used to train the model and optimize the hyperparameters (train set), while the
162 remaining 20% was used to validate the model's performance (test set). Model performance was
163 optimized through hyperparameter tuning. A grid search approach was employed to adjust key
164 hyperparameters, including learning rate and number of trees. To prevent overfitting, fivefold cross-
165 validation was applied to evaluate different hyperparameter combinations (Münch, 2023). The
166 optimal hyperparameter set was selected based on minimizing root mean square error (RMSE),
167 ensuring robust performance on unseen data.

168 Following hyperparameter optimization, model predictions were generated using the test set to assess
169 performance indicators. To establish 95% confidence intervals for predictions, this evaluation process
170 was repeated 100 times using bootstrap sampling. The dataset was randomly partitioned into training
171 and test subsets 100 times, with each bootstrap sample created through random sampling with
172 replacement from the original dataset. This methodology enables comprehensive evaluation of model
173 performance across diverse samples, providing robust estimates of predictive accuracy and associated
174 uncertainty. The resulting 95% confidence intervals offer reliable prediction ranges for each
175 performance indicator (Perrignon et al., 2025).

176 Each model was trained using all the decision variables and informative variables from the database.
177 Its performance was then evaluated using the Root Mean Square Error (RMSE) and the R^2 metric.
178 The RMSE measures the deviation between the predicted values and the actual analytical
179 measurements of the objectives, providing an indication of the model's accuracy. The R^2 metric
180 quantifies the proportion of variance in the response variable (performance indicator) that is explained
181 by the model. A low RMSE and high R^2 value indicate a good model accuracy. Random Forest (RF)
182 achieved the best results, for the RMSE and R^2 , explaining it is selected for the study (Supplementary
183 material, Table S1). RF is a machine learning model built using an ensemble of decision trees, which
184 are built from the collected data. RF algorithm (Breiman, 2001) selects subsets of data and variables
185 for tree construction randomly, to obtain a diversity of trees (Garre et al., 2020; Speiser et al., 2019).
186 Each tree returns a prediction of the output variables and the final prediction is calculated by

187 averaging the predictions of each tree (Koay et al., 2022; Ma and Cheng, 2016). RF and others
188 methods were implemented in Python, using the *scikit-learn* library.

189 2.4 Formulation of the multi-objective problem and constraints

190 Three objectives were defined based on performance indicators defined previously:

- 191 - Minimize the fat content (% w/w) with a minimal value higher than 34% w/w
- 192 - Minimize the moisture content (% w/w) with a minimal value higher than 1.7% w/w
- 193 - Maximize crisp's lightness (%)

194 In order to structure the algorithm's search space and ensure that the results are realistic and
195 applicable, all the constraints associated with the crisps manufacturing process were listed. Equality
196 constraints are parameters that cannot be modified. They refer to the 19 quality attributes of the potato
197 batches. Fryers do not control these parameters that remain fixed during the optimization problem.
198 The process parameters are optimized based on them. Inequality constraints correspond to parameters
199 boundaries defined by minimum and maximum values. The boundaries are the minimum and
200 maximum values of each process variable listed in the dataset. The objectives also have their own
201 boundaries that need to be set. These boundaries delimit the target values to be achieved and are set
202 by the industry. The lower boundary for fat content is set at 34% w/w, and for moisture content at
203 1.7% w/w, which are the minimal values to achieve. The objective is to get as close as possible to
204 these boundaries without going below them. Regarding the lightness, the objective is to maximize
205 this variable. By default, optimization algorithms minimize the indicator; therefore, to maximize it, a
206 negative sign must be placed in front of the value.

207 The multi-objective optimization problem of the potato crisp manufacturing process can be
208 formulated and solved once the boundaries and equality constraints are defined (1):

$$209 \quad \text{Minimize } f(x) = [f_{\text{Fat content}}(x), f_{\text{Moisture content}}(x), -f_{\text{Lightness}}(x)] \quad (1)$$

210 *subject to*

$$211 \quad \text{Fat content} \geq 34 \text{ and Moisture content} \geq 1.7$$

$$212 \quad x_i^{\min} \leq x_i \leq x_i^{\max}, \forall i \in \{1, \dots, n\},$$

213 *with* $x_i = i^{\text{th}}$ *process variable*

214 *and* $n = \text{number of process variables}$

215 *and fixed potato quality attributes*

216 2.5 Presentation of potato scenarios

217 To illustrate the multi-objective approach and understand how to adjust the process based on potato
218 quality, two scenarios corresponding to two potato batches characterized by different quality

219 attributes (Table 1) were set up. The first scenario (Scenario 0) corresponds to a potato batch A. The
220 second (Scenario 1) corresponds to a potato batch B. The quality attributes of the potato batches are
221 representative of potato batches handled by the crisp manufacturing industry. Table 1 do not list all
222 the quality attributes of the potato batches A and B but only the quality attributes having a value other
223 than 0 are indicated. Potato batch A represents the most common scenario encountered in crisp
224 manufacturing, where potatoes meet standard quality requirements with minimal defects. In contrast,
225 potato batch B represents an unusual scenario that occurs when potatoes present significant quality
226 issues, such as appearance defects or suboptimal characteristics. Processing such batches are
227 challenging because they require specific adjustments to the process that can significantly impact
228 product quality and yield if not properly managed. These two contrasting scenarios were chosen to
229 demonstrate the adaptability of the multi-objective optimization approach across a large spectrum of
230 potato quality variations encountered in industrial crisp manufacturing. The 19 variables
231 corresponding to the quality attributes of the two batches of potatoes are fixed as inputs for the
232 optimizations.

233 2.6 Solving the multi-objective problem using Nondominated Sorting Genetic Algorithm II

234 The multi-objective problem was solved by the genetic algorithm Nondominated Sorting Genetic
235 Algorithm II (NSGA-II) proposed by (Deb et al., 2002). NSGA-II is based on non-dominated sorting,
236 designed to solve multi-objective optimization problems with quantitative decision variables. The
237 algorithm is divided into several steps (Figure 3). The first step involves generating randomly an
238 initial population P_0 of size 100 (The population size is a parameter that can be modified by the user.
239 Here, the size has been arbitrarily set to 100). Then, a child population (Q_0), also of size 100, is
240 generated from the initial population using tournament selection, crossover (to combine solutions and
241 explore new possibilities) and mutation (to introduce small random changes and maintain diversity).
242 The new population P_1 is then selected by comparing P_0 and Q_0 through non-dominated sorting,
243 which ranks solutions based on their dominance, selecting the solutions that improve the performance
244 of all objectives and by comparing the crowding distances between solutions. These steps are repeated
245 1000 times. This value corresponds to the number of generations of the algorithm and is set by the
246 user. The tests conducted in this study showed that the results remained unchanged for a number of
247 generations higher than 1000. This method was implemented in Python using the *pymoo* library. The
248 algorithm returns a set of 100 acceptable solutions obtained after 1000 iterations.

249 2.7 Identification of the best solution using Technique for Order of Preference by Similarity 250 to Ideal Solution (TOPSIS)

251 To select a solution from the set of solutions proposed by the algorithm, multi-criteria decision support
252 methods are required. Several methods exist in this family, including AHP (Analytic Hierarchy
253 Process), ELECTRE (ELimination and Choice Translating REality), and PROMETHEE (Preference
254 Ranking Organization METHod for Enrichment Evaluation). In this study, the TOPSIS method
255 developed by Hwang and Yoon (1981) was chosen for its ability to rank solutions efficiently and its
256 particular suitability to contexts where objectives are numerous and heterogeneous. TOPSIS is also
257 easy to implement, well documented and available in several computer languages.

258 The selected Pareto-optimal solution is the one with the greatest Euclidean distance from the negative
259 ideal solution and simultaneously the smallest Euclidean distance from the positive ideal solution
260 (Supplementary material, Figure S2). The TOPSIS method allows for indicator weighting, meaning
261 different weights may be assigned to indicators when they hold different economic or regulatory
262 importance. In this study, equal weights were assigned to all objectives in collaboration with industrial
263 experts to ensure a balanced evaluation of all performance indicators. This selected solution provides
264 the best trade-off between all objectives. The method was implemented in Python using the *topsispy*
265 library.

266 3 Results

267 3.1 Overview of the decision variables in the dataset

268 The 9 decision variables are the process parameters involved in crisp quality optimization. They are
269 adjusted directly by the fryers to influence the process flow. Figure 4 illustrates the frequency
270 distribution of the decision variables observed in the database that corresponds to the different levels
271 and frequency of adjustment of these variables in the process by the fryers. The frequency distribution
272 of the variables Peeling time, Slat thickness, Water supply, and Time Belt 1, indicates that some levels
273 are preferentially used by the fryers while other levels have never been tested during the crisp
274 manufacturing process. For Water supply, one level is mainly applied, while four other levels have
275 been tested sporadically. The frequency distribution of slat thickness shows some untested values due
276 to constraints related to the design of the cutting head, which limits the number of available options
277 for this variable. This information will need to be considered in the constraints of the optimization
278 step, and the obtained values should be assessed in light of the technical capabilities of the process.
279 For the variables Elevator drive sprocket frequency, Fryer temperature, Time Belt 2 and Belt 3 drive
280 sprocket frequency, large ranges of levels have been explored indicating that the fryers preferentially

281 tune these variables to control the quality of the crisps. In summary, figure 4 shows that both
282 operational preferences and technical constraints influence the setting of the decision variables. These
283 results highlight the interest of using a multi-objective optimization approach to assess whether the
284 levels implemented for the decision variables are optimal or not with regard to the crisp quality
285 criteria. Indeed, the algorithm can suggest new combination of levels for the decision variables,
286 provided they fall within the defined boundaries.

287 3.2 Evaluation of model performance

288 For the modeling of each crisp physicochemical indicator, 80% of the dataset, representing 120,127
289 individuals, is used to train and optimize the model's hyperparameters. The model of each crisp
290 physicochemical indicator is then validated on the test set to assess its performance using the RMSE
291 and R^2 evaluation metric. The performance of the model for each crisp physicochemical indicator
292 using RF is shown in Figure 5. It shows the predicted values from the trained model as a function of
293 the observed values from the test set as well as the evaluation metrics (RMSE and R^2) for each
294 indicator.

295 The RMSE is expressed in the same unit than the indicator. This metric simply provides an indication
296 of the quality of the approximation. The closer the value is to 0, the better the model's approximation.
297 RMSE result shows that the predictions of the indicator “fat content” are off by on average 0.46%
298 from the observed values. RMSE result shows that, on average, the predictions of the indicator
299 “moisture content” are off by approximately 0.09 from the observed values and by approximately
300 0.40 for the “lightness” indicator. To compare the performance of the models, the R^2 was examined.
301 The predictive model for crisp “fat content” achieved a R^2 of 0.96, indicating an excellent predictive
302 capability. Similarly, the model for “lightness” reached an R^2 of 0.94, also reflecting very good
303 performance. In contrast, the model predicting “moisture content” yielded an R^2 of 0.71, which
304 demonstrates lower performance compared to the other two models. Nevertheless, all these three
305 models were then used for the optimization step.

306 3.3 Comparison of objectives achieved after optimization and best solution selection

307 For the two potato batches, 100 solutions that satisfy the crisp quality objectives were obtained. The
308 Pareto front for the three crisp quality objectives after processing potato batch A and B are represented
309 in Figure 6 and 7, respectively. Each point represents a solution. “Fat content” (%) is represented on
310 the X-axis, “Moisture content” (%) on the Y-axis, and “Lightness” (%) is represented by the lightness
311 gradient. Both Pareto fronts have the same profile. The optimization algorithm successfully obtained
312 solutions close to industrial targets for fat content and moisture. Indeed, fat values ranged between

313 34% and 34.25%, which remains very close to the set target (Figure 6). For moisture, the results fell
314 between 1.7% and 1.8%, with a target of 1.7%, demonstrating excellent precision while fully meeting
315 industrial requirements (Figure 6).

316 Regarding lightness, values ranged from 64% to 70% (Figure 7), with the majority of the solutions
317 clustering around 69%. Since the aim was to maximize this indicator, and the average value was
318 approximately 65% in the database, these results appear highly satisfactory. Solutions with lower
319 lightness values were mainly associated with a fat content of 34% (Figure 7) suggesting a correlation
320 between decreased fat content and reduced lightness demonstrating the algorithm effectively balances
321 the three performance criteria. With targets met for fat content and moisture content, and consistently
322 high lightness values, the proposed solutions can be considered fully satisfactory for industrial
323 applications.

324 Optimization enables the identification of a set of solutions that balances trade-offs between the three
325 performance indicators. The aim is then to select the best compromise from these solutions based on
326 the industry's guidelines. For this purpose, the TOPSIS method was applied, assigning equal weight
327 to each objective to avoid bias. The solution obtained using the TOPSIS method is highlighted in red
328 in Figure 6 and 7. For Scenario 0 (potato batch A), the selected solution for the crisp physicochemical
329 indicators is a fat content of 34 %, a moisture content of 1.7% and a lightness of 70%. For Scenario
330 1 (potato batch B), the selected solution is for a fat content of 34%, a moisture content of 1.7%. and
331 a lightness of 69.8%. The obtained solutions can then be compared to those actually achieved in the
332 industry, as both batches of potatoes have already been produced. The results are presented in Table
333 2. The objectives obtained after optimization are closer to the target value than those achieved in the
334 industry. Indeed, the industrial targets set for fat content and moisture content are met through
335 optimization, whereas in the industry, the values are close but do not exactly match the target. For
336 lightness, the results after optimization are close to 70%, while in the industry, they are 66.4% for
337 potato batch A and 67.2% for potato batch B. Comparing the industry results with those obtained
338 through optimization demonstrates that the optimization identifies solutions that lead to improved
339 objective values.

340 All these results for the two potato batches indicate that regardless of the quality of the potatoes, the
341 optimization algorithm succeeds in achieving similar objectives. This suggests that the variability in
342 potato quality is corrected during the process, enabling comparable quality objectives for the crisps
343 to be reached. To confirm this hypothesis, it is necessary to closely examine the values of the decision
344 variables for each potato quality.

345 3.4 Scenario comparison and decision support for fryers

346 The methodology integrated into the DT (modeling and optimization) provides a set of solutions for
347 achieving the target crisp quality objectives from a potato batch characterized by specific quality
348 attributes. The set of solutions is a guide for the fryers to better manage their crisp manufacturing
349 lines, by providing the level of the decision variables to be implemented in order to optimize the
350 manufacturing process. The results for the two scenarios are displayed in Figure 8 in the order of
351 appearance of the decision variables in the process. The distribution of the process parameter levels
352 of the database is represented with boxplot and the red circles and the blue triangles indicate the levels
353 selected by the TOPSIS method. Red circles correspond to the solution for scenario 0 (potato batch
354 A), while blue triangles represent the solution for scenario 1 (potato batch A). This figure shows the
355 levels returned after optimization in the light of the range of levels set in the industry, as well as the
356 difference between the levels obtained after optimization for the two scenarios.

357 For the variables “Elevator drive sprocket frequency”, “Slat Thickness”, “Boiler Temperature”,
358 “Fryer Temperature” and “Time Belt 1”, the levels of the process parameters obtained after
359 optimization lie within the distribution of levels already set by the fryers. This means that these levels
360 have already been set for crisp manufacture and, in the opinion of the fryer, do not pose a problem
361 for the correct operation of the process. Conversely, the optimization returns levels outside the
362 distribution of levels observed in the plant for the process parameters “Peeling Time”, “Water
363 Supply”, “Time Belt 2” and “Belt 3 drive sprocket frequency”. In order to assess the relevance of
364 these levels, the opinion of the fryer is required. For "Peeling Time," the result for Scenario 0 (potato
365 batch A) falls within the range of values already used in the industry. For Scenario 1 (potato batch B),
366 the peeling time is nearly twice the mean value used in the industry but remains feasible and
367 acceptable for production. For the “Water Supply” variable, the level obtained after optimization for
368 Scenario 1 (potato batch B) is lower than those set in the industry. This suggests that the operator
369 could reduce by a factor of near 4 the mean amount of water used to wash the potato slats. This
370 solution does not appear to be out of place and would be advantageous for the company from an
371 economic and environmental point of view. With regard to time spent on belt 2, the levels obtained
372 after optimization suggest an increase of “Time Belt 2”, mainly for Scenario 0. Increasing the time
373 spent in the fryer means longer cooking times which in turn improves crisp quality. Finally, for the
374 variable “Belt 3 drive sprocket frequency”, the level obtained after optimization suggests an increase
375 of the speed. The level of this parameter, which influences the fat content of potato crisps, is currently
376 modified within a restricted range of levels by operators. The level obtained after optimization could
377 be of interest as this variable impacts the crisp quality objectives, but in-plant trials are necessary to
378 validate the new setting of the variable “Belt 3 drive sprocket frequency”.

379 The level of decision variables “Slat Thickness”, “Boiler Temperature” and “Fryer temperature”
380 remained unchanged between Scenario 0 and Scenario 1, indicating that the quality differences
381 between the potato batches A and B do not require any adjustment of these parameter settings. In
382 contrast, the level of the variables “Water supply”, “Time Belt 1” and “Time Belt 2” is higher in
383 scenario 0 (potato batch A). This indicates that for potato batch A (without appearance defect), both
384 the conveyor belt times at the fryer inlet and the water supply should be increased. While the
385 difference for “Time Belt 1” between the two scenarios is minimal, “Time Belt 2” requires a
386 significant reduction of 30 s for potato batch B (potato with appearance defects). This indicates that
387 when potatoes have less defects, the potato slats need to spend more time at the second belt in fryer,
388 which is achieved by slowing down the speed of belt 2. For “Water supply”, a higher volume of water
389 is used to wash potato slats for Scenario 0 (potato batch A) compared to Scenario 1 (potato batch B)
390 indicating that slats from potatoes with less appearance defects require more water to meet crisp
391 quality objectives. This variation may also be related to the potatoes size distribution between the 2
392 batches. Indeed, potato batch A contains a higher proportion of smaller potatoes (Table 1), coming
393 with smaller and more numerous slats, which could affect the amount of water needed for the potato
394 slat’s washing. This insight demonstrates that quality of the potato batch can affect the water usage.
395 The level of the process parameters “Peeling Time”, “Elevator drive sprocket frequency” and “Belt 3
396 drive sprocket frequency” should be increased for potato batch B (Scenario 1) compared to potato
397 batch A (Scenario 0). The peeling time of potatoes having defects such as blemishes, deformities, or
398 green spots, *etc.* must be longer than for potatoes without defects in order to obtain uniformly
399 peeled potatoes with no more traces of defects. The speed of the elevator determines the process
400 throughput because this constitutes the limiting step of the crisp manufacturing process. However,
401 the difference of Elevator drive sprocket frequency between the two potato batches remains small.
402 The increase of the speed of the final conveyor belt in the fryer (variable “Belt 3 drive sprocket
403 frequency”) is higher for potato batch B, meaning the potato slats spent less time in the final section
404 of the fryer (Belt 3) for having the same oil content.

405 As with the objectives in the previous section, the values of the decision variables implemented in
406 the industry for these potato batches can be compared with those suggested by the optimization. This
407 comparison helps to understand the decisions made by the fryers and identify necessary adjustments
408 (Supplementary material, Figure S3).

409 **4 Discussion**

410 The illustration of the methodology applied to two batches of potatoes having different qualities
411 reflects the day-to-day reality of food industry processing raw materials with quality attributes that

412 varies with phenotypes, seasons, *etc* (Yee and Bussell, 2007). To overcome this variability, the food
413 industries have two strategies, depending on the raw material being processed. The first one is to
414 standardize the raw material, as many dairy industries do, in order to process a milk of constant
415 composition under constant processing conditions all year round (Soodam and Guinee, 2018). The
416 second strategy is to adapt the process according to the variability of the raw material since it cannot
417 be standardized (Bourquard, 2018). In the field of crisp production, fryers constantly adjust process
418 parameters to variabilities in the quality of potato batch while maintaining regular objectives in terms
419 of crisp qualities. This requires a long training and expertise for the fryers on their crisp manufacturing
420 lines. The methodology developed in this study offers the possibility of adapting to this variability by
421 training machine learning models on a dataset collected over several years. The results obtained for
422 both potato batches demonstrate the necessity of adapting the process based on raw material quality
423 to achieve similar objectives. Table S2 in the supplementary material summarizes the different
424 adjustments to be made on the production line along with the associated expert evaluation for each
425 proposed modification. This expert validation proves essential since the developed tool aims to
426 provide decision support to fryer operators. The acceptance and adoption of this approach by end
427 users directly conditions its industrial success; it is therefore crucial to integrate their feedback to
428 validate and refine the methodology, particularly by adjusting variable bounds or process-related
429 constraints. For example, regarding the “Water supply” variable, the recommendation for one
430 scenario is significantly lower than current industrial practice. While this recommendation makes it
431 possible to reduce crisp line water consumption, it remains questionable as this drastic reduction could
432 impact performance indicators not yet integrated in the optimization. This observation raises several
433 improvement perspectives for the method, such as the integration of additional performance
434 indicators in the optimization or the revision of bounds to better reflect actual operational constraints.
435 Models presented in this article have been studied in two ways. A training phase, on the whole
436 database, ensures its genericity and adaptability to new data (evaluated through a cross-validation).
437 Thanks to the richness of the historical database (two complete years of process between 2022 and
438 2023), the trained models are thus easily transferable to new applications. Retraining them would
439 only be necessary in case of an evolution of the raw material (different sizes, shapes, ...), causing a
440 drift between the predicted and actual values. A validation phase has been done on two distinct batches
441 in order to illustrate two different real-life situations. In a first case, we consider a batch of standard
442 quality potatoes (batch A), without any defect, typical of those processed daily by the fryers. In a
443 second time, an extreme scenario has been processed (batch B) to test the limit of the predictive
444 models. In this case, the potatoes presented many defects in order to represent the most difficult cases

445 to process. While these two batches do not cover the whole range of possible scenarios, they allow
446 us to demonstrate the adaptability of our models in a context of a DT approach.

447 The quality of the models is crucial for the subsequent steps of the multi-objective optimization
448 approach (Madoumier et al., 2019). Indeed, the results of the optimization will be biased and
449 unrealistic if a model does not predict the indicator correctly or if its prediction error is too high. In
450 the context of crisp production, it was shown above that the performance of the "moisture content"
451 model was lower than the two other models, with an R^2 of 0.71 compared to 0.96 and 0.94 for fat
452 content and lightness models, respectively. While this model remains satisfactory given its overall
453 performance metrics (RMSE and R^2) and can be used for the optimization step, its lower prediction
454 accuracy introduces uncertainty into one of the three objectives. This limitation could potentially
455 affect the reliability of optimization outcomes. The reduced model performance may lead to
456 suboptimal solutions or increased variability in the recommended process parameters. Beyond model
457 accuracy, understanding model interpretability is also essential for building confidence in the
458 optimization results. Machine learning models often operate as "black boxes" with numerous
459 interdependent parameters, limiting the interpretation of results and understanding of the model's
460 decision (Przybył, 2024). There are several methods to help interpret models (Ali et al., 2023),
461 including the SHapley Additive exPlanations (SHAP) (Lundberg et al., 2018) and Local Interpretable
462 Model-agnostic Explanations (LIME) (Ribeiro et al., 2016) methods. Both model-agnostic methods
463 help explain how models make their predictions and improve understanding of model behavior.
464 Understanding how this model uncertainty propagates through the optimization process and impacts
465 the final optimal solutions represents an important area for future investigations. Methods such as
466 uncertainty quantification, sensitivity analysis of optimization results to model prediction errors and
467 model interpretability methods such as SHAP and LIME analysis could provide valuable insight into
468 the reliability of the methodology and strengthen user confidence in the optimization
469 recommendations.

470 Another aspect to consider is the computational cost associated with this methodology. The training
471 phase of the models is particularly demanding in terms of computing time (few hours, depending on
472 the amount of training data). Unlike mechanistic modeling based on known equations, data-driven
473 modeling requires a preliminary training phase before using the models for the optimization step
474 (Khan et al., 2022). This training process must be conducted in advance to ensure real-time results.
475 In the case of potato crisp production, the model training phase (including hyperparameter
476 optimization) requires 1 hours on a dedicated server, while the optimization step to determine the
477 parameters to implement for a given potato batch takes only a few seconds (less than 40 in our case).
478 This rapid optimization enables real-time actions to be taken on the process: as soon as a new potato

479 batch is qualified, fryer operators instantly obtain the optimal parameters to be implemented and they
480 control the line according to these recommendations. However, prior model training presents a
481 limitation: once trained, models cannot integrate new parameters or updated data that were not present
482 in the initial training dataset. In this case, models can be retrained to refine future predictions. To
483 overcome this limitation, dynamic model training could be directly integrated into the digital twin.
484 This approach would enable continuous model improvement by incorporating new data collected
485 from the crisps line (Kreuzer et al., 2024). This dynamic training can be performed in parallel without
486 interfering with ongoing optimization processes, allowing newly trained models to be integrated into
487 the DT framework once ready. However, integrating dynamic training requires careful management
488 of computational resources and timing. A practical solution involves implementing a scheduled
489 update framework (whether daily, weekly, or monthly) to systematically retrain models and deploy
490 updated versions in industrial operations (Metcalf et al., 2025).

491 The computing time of the optimization step is directly influenced by two main parameters: the
492 number of generations and the population size used in the algorithm. A high number of generations
493 ensures a high-quality Pareto front but also results in significant computational costs, sometimes for
494 solutions that could be obtained for a lower number of generations. Conversely, if the number of
495 generations is too low, the Pareto front may not be stabilized, leading to suboptimal solutions.
496 Therefore, fine-tuning these parameters is necessary to maximize optimization efficiency while
497 minimizing computational cost (Zakaria et al., 2011). The combination of parameters is specific to
498 each problem. Indeed, depending on the number of objectives to be optimized, the number of
499 generations will need to be optimized to ensure the quality of the Pareto front (di Pierro et al., 2007).
500 In the case of the crisp process optimization, the number of generations was set to 1000 and the
501 population size to 100. This combination of parameters was sufficient in this case study to achieve a
502 stable Pareto front, but it could be optimized to reduce computation time. Currently, the computation
503 time for optimization is short enough to allow for the consideration of potato quality and to send the
504 information back to the fryers. Thus, while the developed methodology demonstrates strong
505 adaptability and generalization potential, its computational cost must be carefully considered,
506 particularly for efficient industrial implementation. The integration of dynamic training strategies and
507 the optimization of the trade-off between computing time and solution quality represent promising
508 outlook for future improving this methodology.

509 To sum up, several perspectives emerge from this work to further enhance the proposed optimization
510 framework. First, the generalization of this data-driven methodology to other food manufacturing
511 processes represents a promising avenue (Datta et al., 2022; Mavani et al., 2022), though success will
512 inherently depend on the quality and representativeness of the available industrial datasets. The

513 adaptability of the approach across different production contexts will require careful consideration of
514 data characteristics, process specificities, and the nature of process parameters (continuous vs.
515 discrete variables), which is crucial for selecting the appropriate optimization methodology. Second,
516 integrating uncertainty quantification into both the machine learning models and optimization process
517 constitutes a stimulating research direction. Addressing model prediction uncertainties and their
518 propagation through the optimization algorithm would provide more robust and reliable solutions for
519 industrial decision-making. Third, industrial deployment and validation of the process parameter
520 predicted by the algorithm remains essential for establishing its practical value. To date, the proposed
521 methodology operates in offline mode, meaning it is not applied in real-time on production lines.
522 However, the definition of a digital twin states that these systems run in parallel to the real food
523 manufacturing process through a continuous data flow. For our approach to be fully qualified as a
524 digital twin, it must be implemented in the industry to provide real-time recommendations.

525 **5 Conclusion**

526 The study presented in this article demonstrates the potential of the digital twin, coupled with a multi-
527 objective optimization approach, as a tool to assist fryers in managing potato crisp production lines.
528 By exchanging a continuous flow of information, the tool integrates collected data throughout the
529 crisp manufacturing process to build models, optimize process parameters, and provide instructions
530 to the fryers to reach crisp quality objectives. The results obtained for potato batches with two
531 different qualities show that the production line needs to be managed differently depending on the
532 quality of the potatoes. This validates the relevance of the present approach in providing real-time
533 specifications on the process parameter settings to be implemented to guarantee optimal crisp quality
534 objectives.

535 Once implemented, it will be able to process real-time data on potato quality and return instructions
536 to the fryers on the level of process parameters to control potato crisp manufacturing process. One of
537 the advantages of the proposed approach is its modularity and its ability to evolve in line with
538 industry's guidelines. While the current optimization focused on three objectives, future
539 developments represent a key opportunity to incorporate environmental sustainability metrics,
540 including water consumption, energy usage, and waste reduction, thereby enabling optimization
541 frameworks that simultaneously address economic and environmental performances. Consequently,
542 the development of digital twins holds promise for addressing current industrial challenges while
543 anticipating future needs, particularly in terms of environmental impact.

544 Funding sources

545 This project was supported by Fondation Institut Agro

546 Acknowledgements

547 The authors thank Romuald Rituit, Laurent Mollard and Geoffrey Ruiz from Altho for sharing the
548 data and for their expertise in interpreting the results. The authors would also like to thank Manon
549 Gaudin, Océane Frayssinet and Nolwenn Poulain for their assistance in collecting and pre-
550 preprocessing the data.

551 Author contributions: CRediT

552 Manon Perrignon: Conceptualization, Investigation, Methodology, Visualization, Writing – original
553 draft; Mathieu Emily: Supervision, Conceptualization, Methodology, Writing – review and editing;
554 Mélanie Munch: Methodology, Writing – review and editing; Paul Debuire: Methodology; Romain
555 Jeantet: Supervision, Conceptualization, Project administration, Funding acquisition, Writing –
556 review and editing; Thomas Croguennec: Supervision, Conceptualization, Writing – review and
557 editing

558 Abbreviations

559 DT = Digital Twin

560 RF = Random Forest

561 RMSE = Root Mean Square Error

562 NSGA-II = Nondominated Sorting Genetic Algorithm II

563 TOPSIS = Technique for Order of Preference by Similarity to Ideal Solution

564 Conflicts of Interest

565 The authors declare no conflicts of interest.

566 References

- 567 Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J.M., Confalonieri, R.,
568 Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., Herrera, F., 2023. Explainable Artificial
569 Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial
570 Intelligence. *Information Fusion* 99, 101805. <https://doi.org/10.1016/j.inffus.2023.101805>
- 571 Barricelli, B.R., Casiraghi, E., Fogli, D., 2019. A Survey on Digital Twin: Definitions, Characteristics,
572 Applications, and Design Implications. *IEEE Access* 7, 167653–167671.
573 <https://doi.org/10.1109/ACCESS.2019.2953499>
- 574 Bourquard, B., 2018. Raw Material Variability in Food Manufacturing. Open Access Dissertations.

- 575 Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32.
576 <https://doi.org/10.1023/A:1010933404324>
- 577 Bruno, F., Ledbetter, M., Davies, B., Riedinger, L., Blidi, S., Sturrock, K., McNamara, G., Montague,
578 G., Fiore, A., 2024. Effect of ultrasound and additives treatment as mitigation strategies to
579 reduce acrylamide formation in potato crisps on industrial scale. *LWT* 197, 115876.
580 <https://doi.org/10.1016/j.lwt.2024.115876>
- 581 Datta, A., Nicolai, B., Vitrac, O., Verboven, P., Erdogdu, F., Marra, F., Sarghini, F., Koh, C., 2022.
582 Computer-aided food engineering. *Nat Food* 3, 894–904. <https://doi.org/10.1038/s43016-022-00617-5>
- 583
- 584 Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic
585 algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6, 182–197.
586 <https://doi.org/10.1109/4235.996017>
- 587 di Pierro, F., Khu, S.-T., Savic, D.A., 2007. An Investigation on Preference Order Ranking Scheme
588 for Multiobjective Evolutionary Optimization. *IEEE Transactions on Evolutionary*
589 *Computation* 11, 17–45. <https://doi.org/10.1109/TEVC.2006.876362>
- 590 Drogenik, J., Pahor, B., Kravanja, Z., Pintarič, Z.N., 2023. Multi-objective scenario optimization of
591 the food supply chain – Slovenian case study. *Computers & Chemical Engineering* 172,
592 108197. <https://doi.org/10.1016/j.compchemeng.2023.108197>
- 593 Garre, A., Ruiz, M.C., Hontoria, E., 2020. Application of Machine Learning to support production
594 planning of a food industry in the context of waste generation under uncertainty. *Operations*
595 *Research Perspectives* 7, 100147. <https://doi.org/10.1016/j.orp.2020.100147>
- 596 Glassey, J., Barone, A., Montague, G.A., Sabou, V., 2017. Case Studies in Modelling, Control in Food
597 Processes, in: Hitzmann, B. (Ed.), *Measurement, Modeling and Automation in Advanced*
598 *Food Processing, Advances in Biochemical Engineering/Biotechnology*. Springer
599 International Publishing, Cham, pp. 93–120. https://doi.org/10.1007/10_2017_13
- 600 Gunantara, N., 2018. A review of multi-objective optimization: Methods and its applications. *Cogent*
601 *Engineering* 5, 1502242. <https://doi.org/10.1080/23311916.2018.1502242>
- 602 Henrichs, E., Noack, T., Pinzon Piedrahita, A.M., Salem, M.A., Stolz, J., Krupitzer, C., 2022. Can a
603 Byte Improve Our Bite? An Analysis of Digital Twins in the Food Industry. *Sensors* 22, 115.
604 <https://doi.org/10.3390/s22010115>
- 605 Hwang, C.-L., Yoon, K., 1981. Multiple Attribute Decision Making, *Lecture Notes in Economics and*
606 *Mathematical Systems*. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-48318-9>
- 607
- 608 Khan, Md.I.H., Sablani, S.S., Nayak, R., Gu, Y., 2022. Machine learning-based modeling in food
609 processing applications: State of the art. *Comp Rev Food Sci Food Safe* 21, 1409–1438.
610 <https://doi.org/10.1111/1541-4337.12912>
- 611 Koay, H.V., Chuah, J.H., Chow, C.-O., Chang, Y.-L., 2022. Detecting and recognizing driver
612 distraction through various data modality using machine learning: A review, recent advances,
613 simplified framework and open challenges (2014–2021). *Engineering Applications of*
614 *Artificial Intelligence* 115, 105309. <https://doi.org/10.1016/j.engappai.2022.105309>
- 615 Konak, A., Coit, D.W., Smith, A.E., 2006. Multi-objective optimization using genetic algorithms: A
616 tutorial. *Reliability Engineering & System Safety* 91, 992–1007.
617 <https://doi.org/10.1016/j.ress.2005.11.018>
- 618 Kreuzer, T., Papapetrou, P., Zdravkovic, J., 2024. Artificial intelligence in digital twins—A systematic
619 literature review. *Data & Knowledge Engineering* 151, 102304.
620 <https://doi.org/10.1016/j.datak.2024.102304>
- 621 Kritzing, W., Karner, M., Traar, G., Henjes, J., Sihm, W., 2018. Digital Twin in manufacturing: A
622 categorical literature review and classification. *IFAC-PapersOnLine*, 16th IFAC Symposium
623 on Information Control Problems in Manufacturing INCOM 2018 51, 1016–1022.
624 <https://doi.org/10.1016/j.ifacol.2018.08.474>

- 625 Lundberg, S.M., Erion, G., Lee, S.-I., 2018. Consistent Individualized Feature Attribution for Tree
626 Ensembles. *Computer Science, Mathematics*.
- 627 Ma, J., Cheng, J.C.P., 2016. Identifying the influential features on the regional energy use intensity
628 of residential buildings based on Random Forests. *Applied Energy* 183, 193–201.
629 <https://doi.org/10.1016/j.apenergy.2016.08.096>
- 630 Madoumier, M., Trystram, G., Sébastien, P., Collignan, A., 2019. Towards a holistic approach for
631 multi-objective optimization of food processes: A critical review. *Trends in Food Science &
632 Technology* 86, 1–15. <https://doi.org/10.1016/j.tifs.2019.02.002>
- 633 Martinez, E., Rodriguez, J.A., Mondragon, A.C., Lorenzo, J.M., Santos, E.M., 2019. Influence of
634 Potato Crisps Processing Parameters on Acrylamide Formation and Bioaccessibility.
635 *Molecules* 24, 3827. <https://doi.org/10.3390/molecules24213827>
- 636 Mavani, N.R., Ali, J.M., Othman, S., Hussain, M.A., Hashim, H., Rahman, N.A., 2022. Application
637 of Artificial Intelligence in Food Industry—a Guideline. *Food Eng Rev* 14, 134–175.
638 <https://doi.org/10.1007/s12393-021-09290-z>
- 639 Metcalfe, B., Acosta-Pavas, J.C., Robles-Rodriguez, C.E., Georgakilas, G.K., Dalamagas, T., Aceves-
640 Lara, C.A., Daboussi, F., Koehorst, J.J., Corrales, D.C., 2025. Towards a machine learning
641 operations (MLOps) soft sensor for real-time predictions in industrial-scale fed-batch
642 fermentation. *Computers & Chemical Engineering* 194, 108991.
643 <https://doi.org/10.1016/j.compchemeng.2024.108991>
- 644 Peñalver-Soto, J.L., Muñoz-Guillermo, M., Garre, A., Iguaz, A., Fernández, P.S., Egea, J.A., 2022.
645 Multiobjective Optimization of a Frying Process Balancing Acrylamide Formation and
646 Quality: Solution Analysis and Uncertainty Propagation. *Foods* 11, 3689.
647 <https://doi.org/10.3390/foods11223689>
- 648 Perrignon, M., Croguennec, T., Jeantet, R., Emily, M., 2024. The multi-objective data-driven
649 approach: A route to drive performance optimization in the food industry. *Trends in Food
650 Science & Technology* 152, 104697. <https://doi.org/10.1016/j.tifs.2024.104697>
- 651 Perrignon, M., Emily, M., Munch, M., Jeantet, R., Croguennec, T., 2025. Machine learning for
652 predicting industrial performance: Example of the dry matter content of emmental-type
653 cheese. *International Dairy Journal* 162, 106143.
654 <https://doi.org/10.1016/j.idairyj.2024.106143>
- 655 Przybył, K., 2024. Explainable AI: Machine Learning Interpretation in Blackcurrant Powders.
656 *Sensors* 24, 3198. <https://doi.org/10.3390/s24103198>
- 657 Ribeiro, M.T., Singh, S., Guestrin, C., 2016. “Why Should I Trust You?”: Explaining the Predictions
658 of Any Classifier, in: *Proceedings of the 22nd ACM SIGKDD International Conference on
659 Knowledge Discovery and Data Mining, KDD '16*. Association for Computing Machinery,
660 New York, NY, USA, pp. 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- 661 Salvador, A., Varela, P., Sanz, T., Fiszman, S.M., 2009. Understanding potato chips crispy texture by
662 simultaneous fracture and acoustic measurements, and sensory analysis. *LWT - Food Science
663 and Technology* 42, 763–767. <https://doi.org/10.1016/j.lwt.2008.09.016>
- 664 Shankarrao Patange, G., Bharatkumar Pandya, A., 2022. How artificial intelligence and machine
665 learning assist in industry 4.0 for mechanical engineers. *Materials Today: Proceedings*
666 S221478532205369X. <https://doi.org/10.1016/j.matpr.2022.08.201>
- 667 Soodam, K., Guinee, T.P., 2018. The case for milk protein standardisation using membrane filtration
668 for improving cheese consistency and quality. *Int J of Dairy Tech* 71, 277–291.
669 <https://doi.org/10.1111/1471-0307.12502>
- 670 Speiser, J.L., Miller, M.E., Tooze, J., Ip, E., 2019. A comparison of random forest variable selection
671 methods for classification prediction modeling. *Expert Systems with Applications* 134, 93–
672 101. <https://doi.org/10.1016/j.eswa.2019.05.028>
- 673 Verboven, P., Defraeye, T., Datta, A.K., Nicolai, B., 2020. Digital twins of food process operations:
674 the next step for food process models? *Current Opinion in Food Science* 35, 79–87.
675 <https://doi.org/10.1016/j.cofs.2020.03.002>

- 676 Viklund, G., Mendoza, F., Sjöholm, I., Skog, K., 2007. An experimental set-up for studying
677 acrylamide formation in potato crisps. *LWT - Food Science and Technology* 40, 1066–1071.
678 <https://doi.org/10.1016/j.lwt.2006.07.012>
- 679 Wari, E., Zhu, W., 2016. A survey on metaheuristics for optimization in food manufacturing industry.
680 *Applied Soft Computing* 46, 328–343. <https://doi.org/10.1016/j.asoc.2016.04.034>
- 681 Yee, N.G., Bussell, W.T., 2007. Good Potatoes for Good Potato Crisps, a Review of Current Potato
682 Crisp Quality Control and Manufacture.
- 683 Zakaria, M.Z., Jamaluddin, H., Ahmad, R., Muhaimin, A.H., 2011. Effects of genetic algorithm
684 parameters on multiobjective optimization algorithm applied to system identification
685 problem, in: 2011 Fourth International Conference on Modeling, Simulation and Applied
686 Optimization. Presented at the 2011 Fourth International Conference on Modeling, Simulation
687 and Applied Optimization (ICMSAO), IEEE, Kuala Lumpur, Malaysia, pp. 1–5.
688 <https://doi.org/10.1109/ICMSAO.2011.5775624>
- 689 Zhang, J., Li, J., Fan, L., 2024. Effect of starch granule size on the properties of dough and the oil
690 absorption of fried potato crisps. *International Journal of Biological Macromolecules* 268,
691 131844. <https://doi.org/10.1016/j.ijbiomac.2024.131844>
- 692

Table 1: Presentation of the two potato quality scenarios defined for optimization. Scenario 0 represents a batch of potatoes A without appearance defect, while Scenario 1 represents a batch of potatoes B with appearance defects (burnt, spotted, green).

Scenario	Dry matter (% w/w)	Small size (%)	Medium size (%)	Quality index	Tuber weight (g)	Burnt (%)	Spotted (%)	Green (%)
Scenario 0: potato batch A	21,2	77,31	21,99	7,5	134,06	0	0	0
Scenario 1: potato batch B	20,7	18,16	71,96	6	245,37	1,28	2,17	5,73

Table 1: Comparison of optimized values for TOPSIS-selected solutions with actual industry results for potato batches. The values in green represent the best results for each indicator and potato batch.

The optimization suggests better objectives than actual objectives

		Fat content	Moisture content	Lightness
		(%)	(%)	(%)
Potato batch A	Actual objectives achieved in the industry	34.9	1.8	66.4
	Optimized objectives	34	1.7	70
Potato batch B	Actual objectives achieved in the industry	34.6	1.8	67.2
	Optimized objectives	34	1.7	69.8

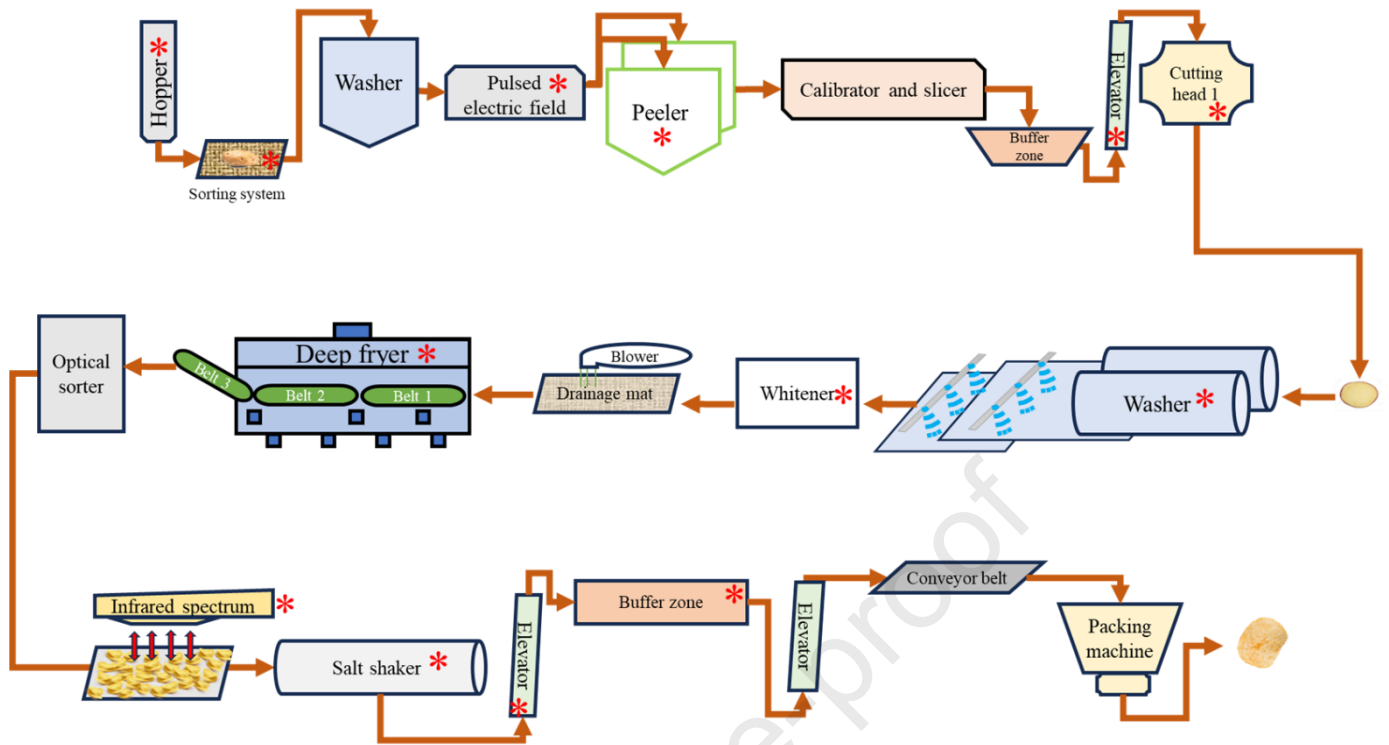


Figure 1: Schematic representation of the potato crisps production. The red stars represent the data collection points.

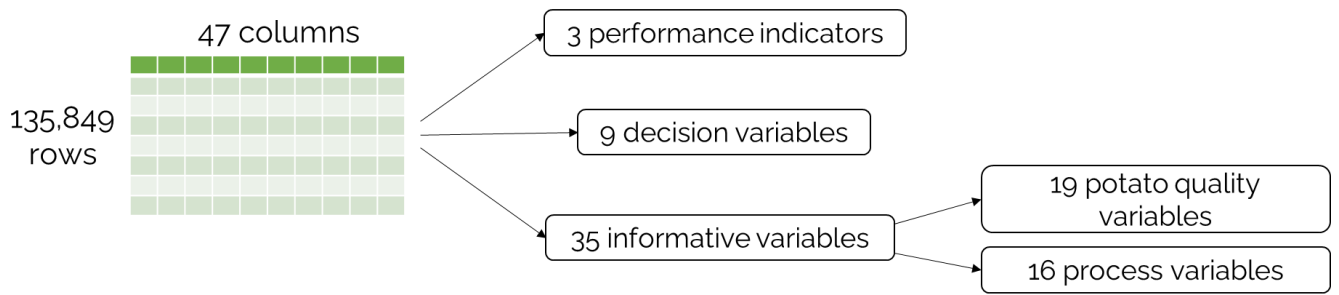


Figure 2: Simplified version of the dataset used in this study and different variable categories. Each column corresponds to a measured process parameter, classified into a specific category (e.g., performance indicators, decision variables and informative variables). Each row represents a batch of potatoes processed in the system. The complete dataset consists of 135,849 rows (individuals) and 47 columns (variables). The informative variables include potato quality variables and process variables.

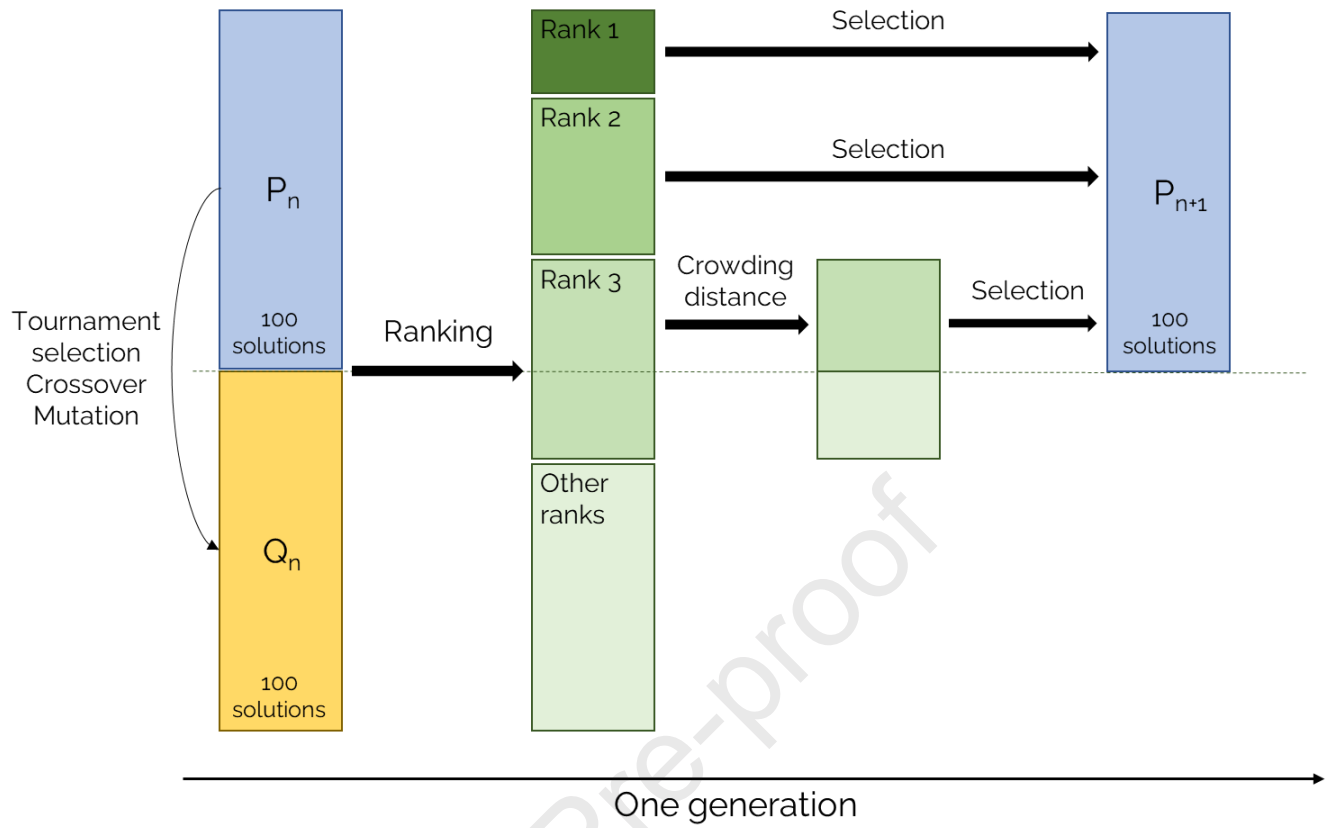


Figure 3: NSGA-II procedure for creating one generation. P_n represents the initial population and Q_n represents the child population generated by P_n through mutation, crossover and tournament selection.

The population size has been set to 100 solutions.

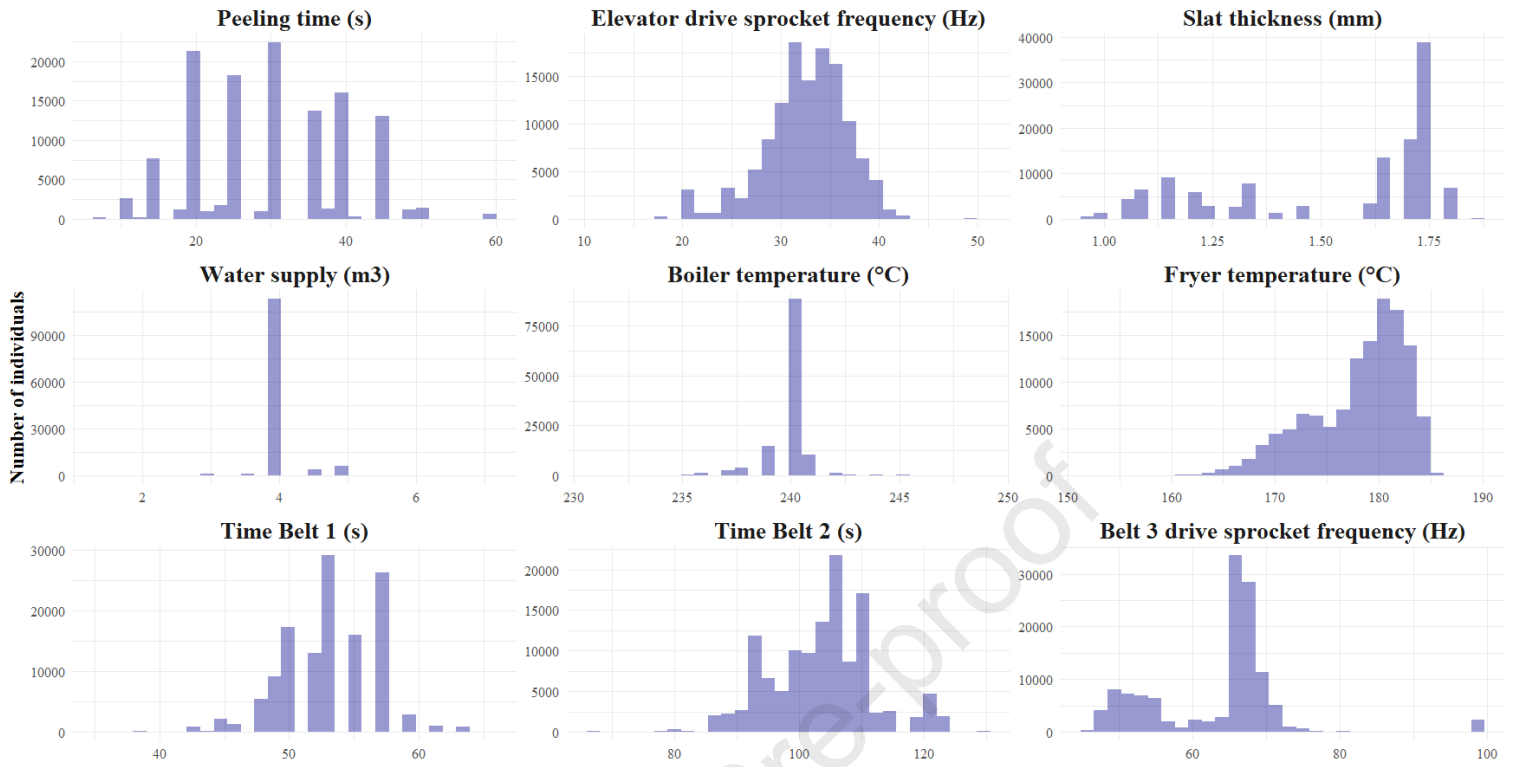


Figure 4: Histograms illustrating the distribution of decision variables observed in the dataset

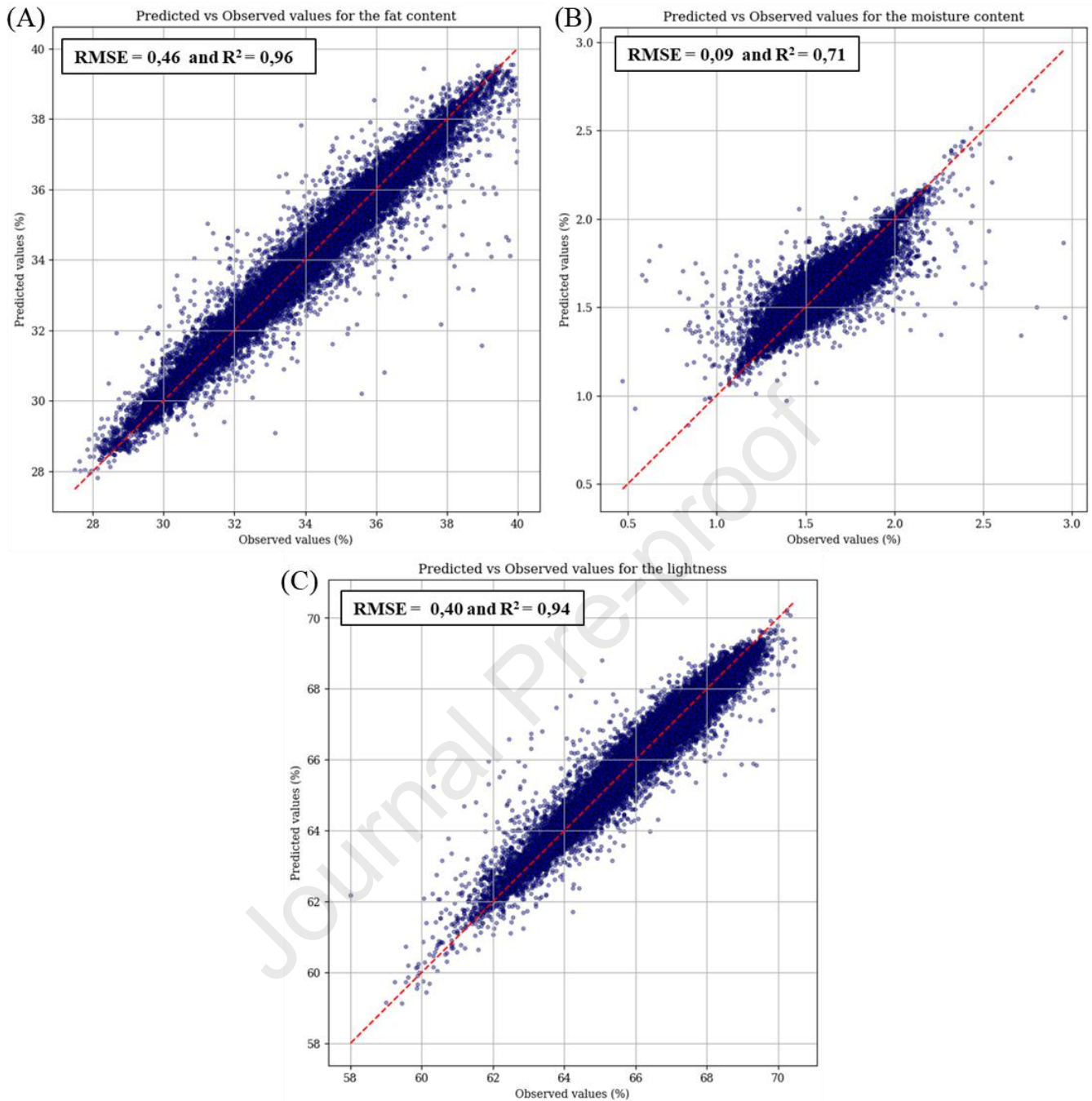


Figure 5: Performance of the three RF models for predicting crisp physicochemical indicators, determined on the test set. Each graph shows the comparison between predicted and observed values, along with performance metrics RMSE (Root Mean Squared Error) and R² (coefficient of determination). Graph (A): Model performance for the *fat content* indicator. Graph (B): Model performance for the *moisture content* indicator. Graph (C): Model performance for the *lightness* indicator.

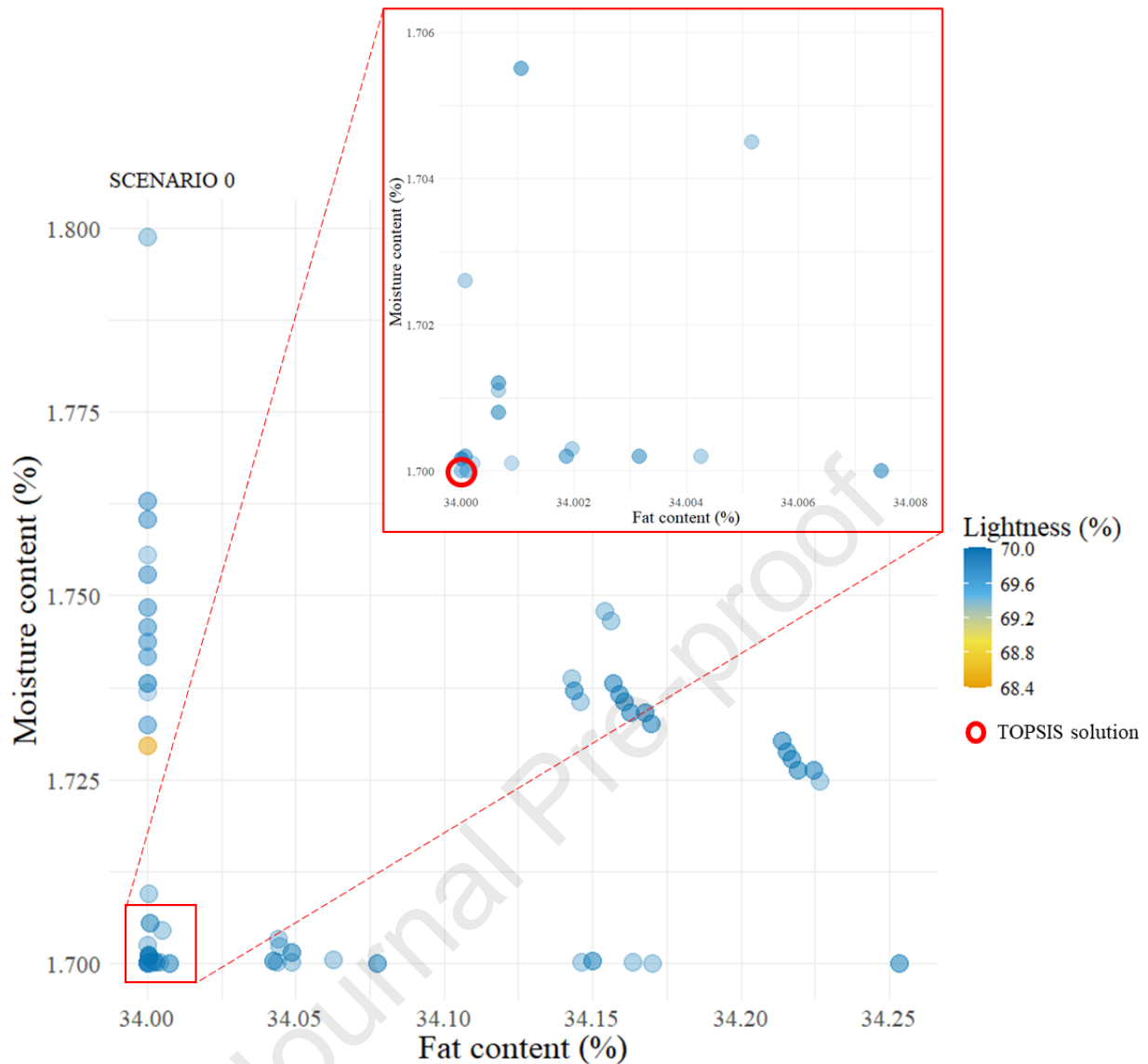


Figure 6: Representation of the Pareto front for the 3 objectives of the potato crisp process after optimization by the NSGAI algorithm (number of generations = 1000 and population size = 100) for Scenario 0 (potato batch A). “Fat content” is represented on the X axis, “Moisture content” is represented on the Y axis and “Lightness” is represented by the lightness scale. The red frame represents a zoomed-in view of the area near the targets for fat and moisture contents. The red circle represents the solution selected from the Pareto front by the TOPSIS method.

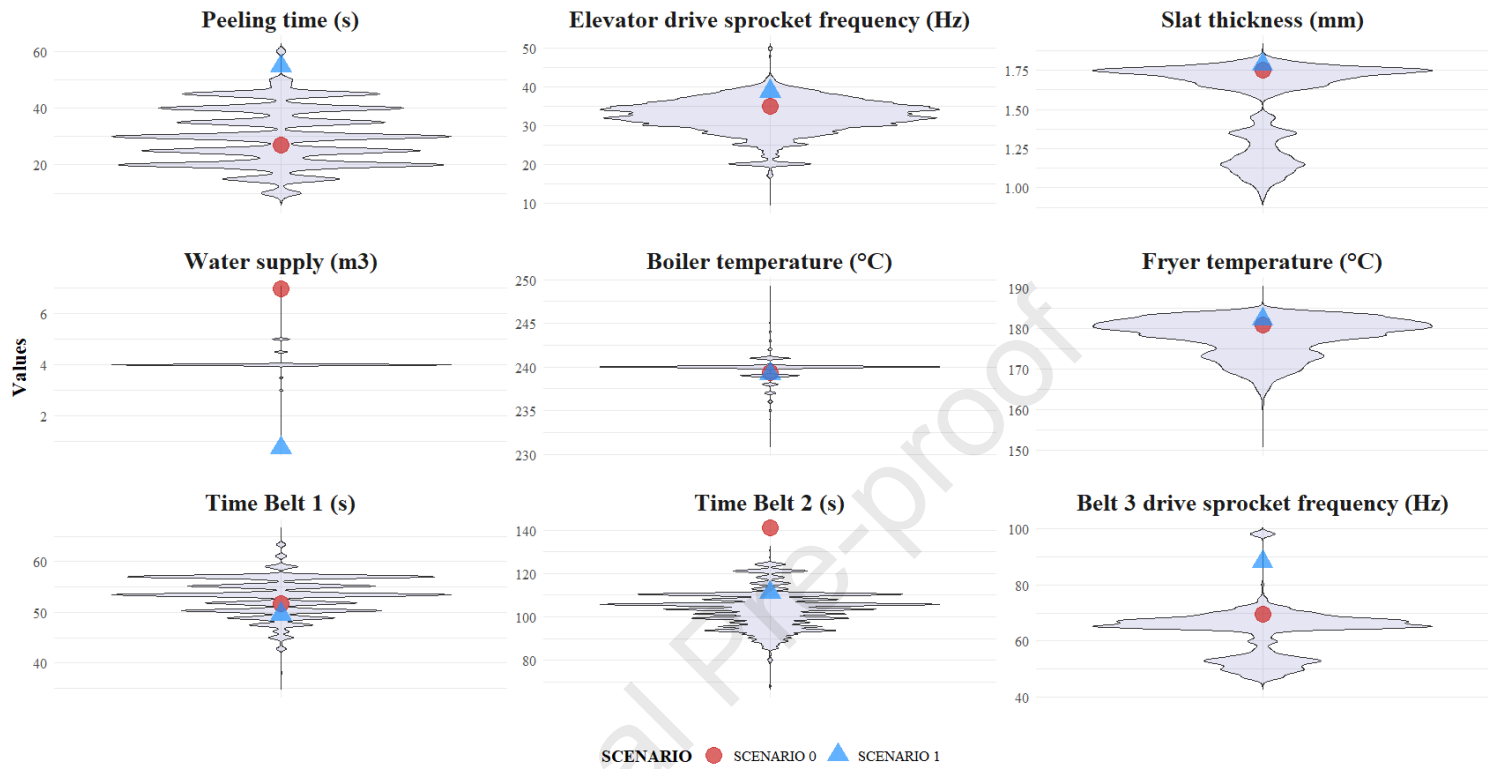


Figure 8: Violin graphs of the process decision variables based on data from the database. Red circles indicate the optimal values obtained for each variable after optimization for scenario 0 (potato batch A), blue triangles represent those for scenario 1 (potato batch B). These optimal values determine the settings to be implemented in the industry to achieve the target objectives.

Highlights:

- Digital twin integrating multi-objective optimization for crisp manufacturing
- Machine learning models predict fat, moisture and lightness performance indicators
- Optimization delivers optimal solutions proving differential control necessity between potato batches
- Process parameters must be adjusted based on potato quality to meet crisp standards
- Framework provides a decision-support tool for process control by fryer operators

Journal Pre-proof

Declaration of interests

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.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof