

Leveraging Causal AI Models for Enhanced Process Control in Food Production: A Case Study on Beer Fermentation Quality

Ž. Kurtanjek*

University of Zagreb Faculty of Food Technology and Biotechnology, Pierrotijeva 6, 10 000 Zagreb, Croatia

This work is licensed under a
Creative Commons Attribution 4.0
International License



Abstract

Reproducible Kernel Hilbert Space (RKHS)-based methodologies are applied for Digital Twins (DT) modelling and causal analysis of the beer fermentation process. The model accounts for the key process variables (temperature, biomass, sugar, ethanol, and carbon dioxide) along with online beer quality monitoring through the concentration of diketones (VDK). The modelling dataset is based on integration of 10 fermentation profiles with the perturbed initial conditions. The focus is on de-confounding process data multicollinearity for development of a DT model with intervention potential aimed for optimising beer quality. The results are graphically presented as a Directed Acyclic Graph (DAG). Based on the DAG structure, a Bayesian Neural Network (BNN) is applied for the graphical presentation of probability density functions of non-linear dependencies of beer quality on the key causal process variables. In addition to the intervention potential of causal DT models, the potential use of counterfactual analysis in causal digital twin model is emphasised for improving process control and possible innovations of new products.

Keywords

BCN Bayes causal network, BIC Bayes information criteria, RKHS reproducible kernel Hilbert space, DAG directed acyclic graph, Digital Twins, beer quality

1 Introduction

Digital Twins (DT) – virtual replication of physical, chemical, and microbial systems – are increasingly being applied in food technology to enhance various aspects of production, processing, quality control, and supply-chain management. Based on the integration of multidimensional, multi-disciplinary big data and Artificial Intelligence (AI), DTs are revolutionising the food industry.^{1–5} The development of AI-based Digital Twins in industrial food production environments must address several challenging issues, some of which are:

- Focusing on applications for interventions in on-line process control and decision-making in production policies.
- Account for stochastic perturbations in the composition and quality of raw feedstocks, as well as instability in supply chains and product markets.
- Accounting for essential time dependencies of process variables.
- Deconfounding multicollinearity among variables with statistically unknown multivariate probability distributions.
- Controlling batch processes with distinct process phases.
- Accounting for the irreversibility of chemical and biochemical states.
- Revealing complex causal inferences related to human perception of food product quality.

Digital Twins, the concept from Industry 4.0, are used to model, simulate, and optimise industrial processes. In the food industry, DTs serve as adaptive, predictive, and control systems that integrate the fields of food engineering and computer science.^{6–7} They propose a hybrid modelling approach that combines traditional food process modelling and simulation with data-driven methods based on machine learning (ML). Research focuses on merging these two fields into an artificial intelligence system for decision-making and process control in stochastic environments. For example, DTs have been proposed for the virtualisation of food supply chains to enhance food safety.^{8–9} Digital Twins have also been developed based on the integration of optical spectroscopy data to infer agro-product quality¹⁰ and beer quality.¹¹ Additionally, a causal Bayesian network has been proposed to account for the complexity and stochastic fluctuations of wine physical and chemical composition on human perception of quality.¹² The aim of this work was to propose a methodology for modelling in Hilbert space using reproducing Kernel Hilbert Space (RKHS) techniques, and the integration of structured causal networks for prediction, probability of intervention effects, and counterfactual evaluation of potential innovations.

2 Methodology

In this work, reproducing kernel Hilbert space (RKHS) is applied for modelling and causal inference. This concept is widely used in machine learning, statistics, and functional analysis. RKHS is a Hilbert space of functions in which evaluation at each point can be represented as an inner scalar product. Its key feature is the existence of a repro-

* Prof. Želimir Kurtanjek, PhD, retired
Email: zelimir.kurtanjek@gmail.com

ducing kernel, which allows for the evaluation of functions in the space using inner products.^{13–14} It greatly simplifies statistical analysis and modelling of non-linear systems by mapping original process features data X from a finite dimensional space R of experimental data into an infinite dimensional Hilbert space $\mathcal{H}$. A kernel function $k(x_1, x_2)$ defines the mapping rule for the self binary product to space of real numbers $k: X \times X \rightarrow R$, and the corresponding feature mapping ϕ of process variables X to Hilbert space $\phi: X \rightarrow \mathcal{H}$. Unlike traditional machine learning algorithms, such as neural networks and decision trees that apply explicit transformation of input data X into feature vectors, kernel methods directly compute similarity using a k user-specified kernel function. For modelling of continuous experimental data, Gaussian weighting (Eq. 1) is commonly applied as the kernel function, (Eq. 1), providing metrics of data similarity:

$$k(x_1, x_2) = \exp\left(-\frac{x_1 - x_2^2}{\sigma^2}\right) \quad (1)$$

Although this kernel function implicitly operates in a high-dimensional, potentially infinite-dimensional feature space, the so-called “kernel trick” of RKHS (Eq. 2), allows to avoid explicitly computing the feature representations. This implicit operation allows kernel methods to efficiently analyse complex data relationships and facilitate effective pattern recognition and analysis in infinite-dimensional Hilbert space $\mathcal{H}$, computation is implemented in finite space R .¹⁴

$$k(x_1, x_2)_R = \langle \phi(x_1), \phi(x_2) \rangle_{\mathcal{H}} \quad (2)$$

Mapping of input data $X(ns, np)$, where ns is the number of samples and np the number of observed variables, yields Gramm matrix $K(ns, ns)$ with the elements

$$K_{i,j} = k(x_i, x_j) \quad (3)$$

Application of Gaussian weights $k(x_i, x_j)$ enables non-parametric auto-regression in Hilbert space, which is effectively applied to data smoothing, removal of measurement errors and improbable outliers. It is applied by use of the weighted interpolation:

$$y_{\mathcal{H}}(x) = \frac{\sum_{i=1}^{ns} k(x, x_i) y_i}{\sum_{i=1}^{ns} k(x, x_i)} \quad (4)$$

For modelling of non-linear regression, a functional relation $f(x)$ based on RKHS representer property is applied:

$$f(x)_R = \langle f, k(\cdot, x) \rangle_{\mathcal{H}} \quad (5)$$

given as linear regression:

$$f(x_i) = \beta_0 + \sum_{j=1}^{j=ns} \beta_j k(x_i, x_j) = \beta_0 + \mathbf{k}^T \boldsymbol{\beta} \quad (6)$$

In the development of Digital Twins as causal AI models, the key focus is on inferring bivariate dependencies and synthesising a directed acyclic graph (DAG) as a network of causal relations. The concept of causal graph, in the form of a Bayesian causal network, is universal, but generally suited to cross-sectional, time-invariant pattern analysis and linear regression, assuming normal distributions of stochastic perturbations.¹⁵ However, processes in the food industry are typically time-dependent, inherently non-linear, and with unknown probability distributions of stochastic disturbances, which need to be accounted for when enabling intervention capability in Digital Twins. RKHS modelling is basically non-linear and non-parametric in a statistical sense. Due to RKHS linear structures, common statistics is applied to the kernel inner-product data. The measure of bivariate relation between X and Y is calculated as the covariance between standardised (zero mean and unity variance) K_x and K_y data, resulting in the Hilbert-Schmidt independence criterion (HSIC). HSIC is a measure of dependence between two process variables X and Y , based on their embeddings in an RKHS.¹⁶ The HSIC coefficient is zero if and only if X and Y are independent.

$$HSIC(X, Y) = \frac{1}{ns^2} \sum_{i=1}^{i=ns} \sum_{j=1}^{j=ns} K_{i,j}(X) K_{i,j}(Y) \quad (7)$$

The direction of bivariate causality for non-collinear X and Y , conditioned by covariates Z , can be inferred based on Kolmogorov's conjecture of the asymmetrical complexity of conditional probabilities between $P(Y|X)$ and $P(X|Y)$.¹⁷ Here, complexity is assessed through a comparison of partial Pearson correlations ($pcor$) between non-parametric predictions of $X \sim X_{\mathcal{H}}(X|Y; Z)$ and $Y \sim Y_{\mathcal{H}}(Y|X; Z)$ using corresponding Gaussian RKHS models:

$$X \rightarrow Y \Rightarrow pcor(Y, Y_{\mathcal{H}}(Y|X; Z)) > pcor(X, X_{\mathcal{H}}(X|Y; Z)) \quad (8)$$

The optimal DAG structure is inferred based on maximum likelihood $\max(L)$ calculated from non-parametric probability density distributions $f_{\mathcal{H}}$, which are evaluated at each DAG_k node using RKHS Gram matrix data. Inference of optimal DAG structure is an ill-conditioned and computationally intensive problem. A critical success factor is the integration of fundamental engineering knowledge with data-driven pattern discovery. In causal AI Digital Twins development, prior knowledge includes structural equation models (SEM) of mass and energy balances, stoichiometric constraints of biochemical reactions, and empirical dependences of mass and energy transfer coefficients and chemical/biochemical rate coefficients. Optimisation of Bayes information criteria (BIC) accounts for DAG complexity determined by the number of model parameters n_p given by the number of network edges, and the number ns of observed samples.

$$L(DAG_k) = \prod_{i=1}^{i=ns} pdf_{\mathcal{H}}(X_i | DAG_k) \quad (9)$$

$$BIC(k) = n_p(k) \log(ns) - 2 \log(L(DAG_k)) \quad (10)$$

In evaluating the likelihood, in addition to the essential role of integrated *a priori* field knowledge, a critical factor is the unknown knowledge of conditional probability distribution (pdfs), which are specific to each individual network node X_i and conditioned on each hypothetical k -th DAG. Application of embedded data into RKHS space enables inference of node-specific non-parametric conditional probability:

$$pdf_{\mathcal{H}}(\widehat{DAG}) = \min_k (BIC(k)) \quad (11)$$

Optimisation of the BIC is itself a numerically demanding problem. Common approaches include heuristic local search methods, such as hill climbing or globally optimal methods like the constrained genetic algorithm GA. Inferred DAG of causal Bayesian network (BN) is the essential part of the Digital Twins models in food production systems (both industrial and agronomic). It enables analysis across all three “rungs of the knowledge ladder”: 1) prediction, 2) intervention, and 3) counterfactual analysis.¹⁵ On the first rung (prediction), the BN provides functionality as standard agnostic (“black box”) machine learning (ML) models, such as neural networks (NN) and decision trees (DT). Their main application is as a part of on-line process control systems. The second “rung” (intervention) allows process control design and/or policy answers to the question of what are the effects $do(X=x)$ of deliberate change or intervention of X on Y when effects of confounding covariates are removed. The third “rung” (counterfactual analysis) provides answers to hypothetical control and process design innovations. Its main objective is to innovate new products, process design and control.

This work adopts directed d-separation methodology, referred to as the $do(X=x)$ calculus, proposed by Pearl.¹⁸ It applies adjustment by structural redesign of BN in the presence of covariates Z and confounding W by modified, “cut off” probability distribution:

$$P(Y \mid do(X=x), Z, W) = P(Y \mid X=x, Z) \quad (12)$$

From the perspective of RKHS-based modelling, unknown probability distributions are inferred from data histograms using non-parametric regression. This work presents graphical depictions of do -calculus probability distributions generated from data by modelling with probabilistic Bayesian neural networks (BNN).

3 Digital twin of beer fermentation

The bio-transformative power of fermentation is fundamentally important in most food production systems. Controlling fermentation parameters influences nutritional quality, flavour and texture, consumer perception of food quality, product preservation, digestive health, and, most importantly, immune system support. From a mathematical and computational perspective, fermentation processes are stochastic, time-dependent (non-stationary), and non-linear systems, making them difficult to model and control.

Digital Twins offer transformative potential in how fermentation processes are managed within the food technology sector, delivering numerous benefits in terms of efficiency, sustainability, and innovation. Digital Twins are information-rich systems that generate big data, enabling real-time process monitoring, precision control, optimisation, minimisation of carbon footprint, predictive maintenance, improved product consistency, innovation of new products, and sustainability by inclusion of fermentation by-products into circular economy. One practical application of a Digital Twin is in energy management during beer production.¹⁹ It is estimated that around 70 % of a brewery's energy demand is spent for heating and cooling fermentation vessels. With the use of an energy Digital Twin for on-line monitoring and control loops, such as heat pumps connected to renewable-energy-powered heat exchangers, significant energy savings can be achieved.

In scientific literature, most studies on fermentation focus on mathematical modelling based on the first principles of mass and heat balances and biochemistry kinetics.^{4,20} These models are reported as systems of ordinary differential equations (ODE) for biomass growth, metabolite concentrations, pH, and temperature. The kinetic parameters are estimated from on-line monitoring laboratory-level beer production systems, and beer quality is monitored from measurement of vicinal diketone concentration (VDK). In this work, a literature-based fermentation model was used to simulate 10 fermentation runs with perturbed initial conditions in the ranges: $T(0) \in [19-28 \text{ }^\circ\text{C}]$, $X(0) \in [0.22-0.28 \text{ g l}^{-1}]$, $S(0) \in [75-85 \text{ g l}^{-1}]$. Typical simulations of fermentation profiles and VDK are shown in Fig. 1. VDKs are formed as by-products during the synthesis of amino acids, valine and isoleucine by yeast, and are produced during the fermentation process, and can impart flavours such as butter, butterscotch, or honey to the beer. The mathematical model integrates fermentation over a 100-h period with hourly sampling. The model (Eq. 13) is a set of six autonomous ODE equations with the state variables: temperature T , yeast concentration X , sugar concentration S , dissolved carbon dioxide concentration CO_2 , ethanol concentration E , concentration of vicinal diketones VDK, and random perturbations (process and measurements noise) $n \in N(0, \sigma)$. The state vector Y is given by:

$$Y = (T, S, X, E, \text{CO}_2, \text{VDK}) \quad (13)$$

$$\frac{d}{dt}Y = f(Y) \quad Y(0) = Y_0 \quad t \in [0, 100]$$

Continuous responses $Y(t)$ are sampled at time intervals $\Delta t = 1 \text{ h}$ with normally distributed noise $n \in N(0, \sigma)$ added. The data for prediction of beer quality by VDK are generated by Eq. (14), and are collected in a data frame $DF = (Y(k), \text{VDK}(k+1))$.

$$\text{VDK}(k+1) = \text{VDK}(k) + \Delta t \cdot f(Y(k)) \quad k = 1, 1000 \quad (14)$$

The DF is Markovian, enabling modelling as a Bayesian network by a directed acyclic graph DAG, and provides the basis for causal analysis.

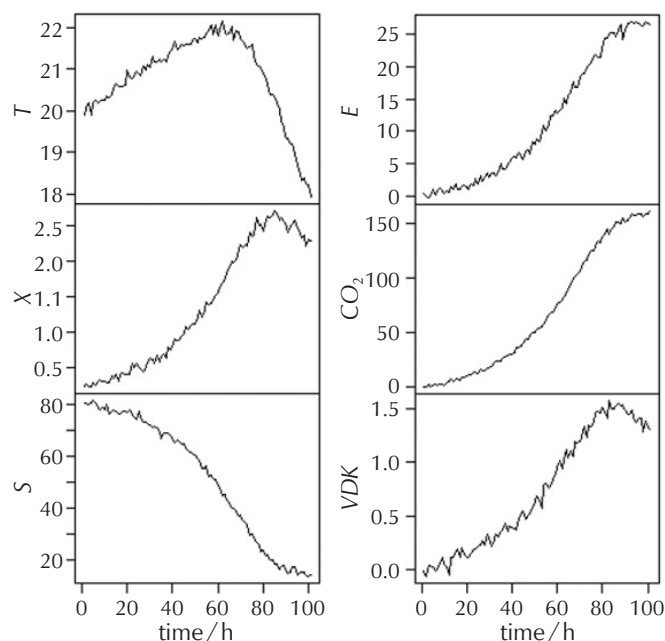


Fig. 1 – Fermentation profiles: time (h), temperature T (°C), biomass X (g l⁻¹), sugar S (g l⁻¹), ethanol E (g l⁻¹), carbon dioxide CO_2 (l), vicinal diketones VDK (ppm)

Slika 1 – Profili fermentacije: vrijeme (h), temperatura T (°C), biomasa X (g l⁻¹), šećer S (g l⁻¹), etanol E (g l⁻¹), ugljikov dioksid CO_2 (l), vicinalni diketon VDK (ppm)

Usually, observed process data are perturbed by random noise, and eventually sporadic outliers. Embedding process data into RKHS enables linear regression to efficiently remove both the random noise component and the improbable outliers via kernel averaging (Eq. 4) and non-parametric regression (Eq. 6). The ridge algorithm developed by Hayfield and Racine *np* code was applied for this purpose.²² In Fig. 2(A), the simulated experimental

VDK concentrations data are shown, along with the corresponding RKHS-smoothed autocorrelation curve. The non-parametrically smoothed quality data are then integrated with the process data. The predictive power of the source data in relation to product quality was tested using machine learning (ML) with random forest (RF) *xgboost* algorithm.²³⁻²⁶ The accuracy of the RF model is presented in Fig. 2(B). The data represent ten fermentation runs, comprising 1,000 samples in total, with 800 samples used for model training and 200 retained untrained for testing. The RF model showed very high accuracy, with a standard error of 2 % for the training set and 3 % for the test set. This very high level of accuracy of the RF model is attributed to the strong correlation based on ensembled non-linear decision trees. Hence, the RF model interpolation accuracy makes it highly applicable in on-line control loop of beer production, but it does not reveal causal relations between process variables and the product quality required for process optimisation and innovations.²⁷⁻²⁹

4 Results and discussion

The initial steps in causal analysis of time-series data from multidimensional non-linear dynamic processes involve testing binary dependencies. The Pearson correlation assumes linear and stationary data, but to account for non-linearity, the process data must be embedded into RKHS high-dimensional space where linear relationships can be more easily inferred and HSIC (Hilbert-Schmidt independence criterion) applied, Eq. (7). Table 1 presents the evaluated HSIC coefficients between the process data ($X = (T, S, X, E, CO_2)$ input variables) and product quality ($Y = VDK$ concentration is output variable). Effect of biomass on VDK has the highest HSIC coefficient, although it is not significantly different from the coefficients for sugar S , ethanol E , and carbon dioxide CO_2 . These binary HSIC values do not account for confounding of the covariates

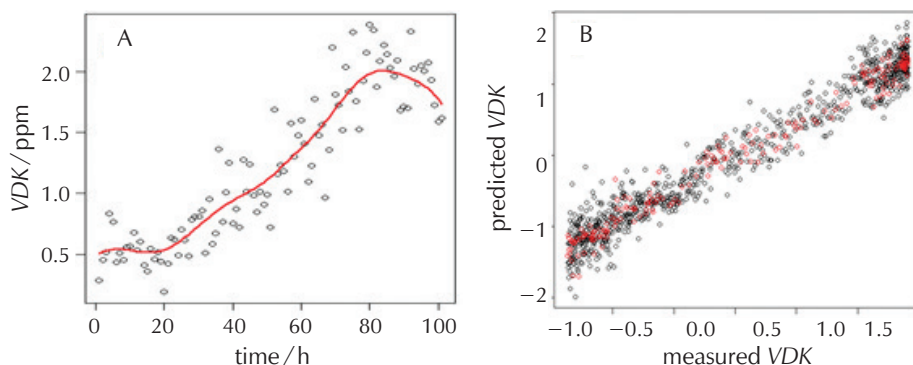


Fig. 2 – Model of VDK vicinal diketones concentration during fermentation, A) experimental data (circles) and RKHS representation (red curve), B) predictions of VDK standardised concentrations (black circles are trained data, red circles are untrained data) by a random forest (RF) model

Slika 2 – Model koncentracije VDK vicinalnih diketona tijekom fermentacije, A) eksperimentalni podatci (krugovi) i RKHS prikaz (crvena krivulja), B) predviđanja VDK standardiziranih koncentracija (crni krugovi su trenirani podatci, crveni krugovi su netrenirani podatci) pomoću modela slučajne šume (RF).

Table 1 – Binary directional causal effects of the process parameters on VDK vicinal diketones. Evaluated are Hilbert-Schmidt independence criteria (HSIC), partial correlations $pcor(X_{\mathcal{H}}, Y_{\mathcal{H}})$, and confidence intervals (C.I.) between cause X and effect Y in Hilbert space, $pcor(Y, Y_{\mathcal{H}})$ partial correlations between the effect Y and its prediction $Y_{\mathcal{H}}$ conditioned on X and the corresponding interval C.I., $pcor(X, X_{\mathcal{H}} | Y)$ partial correlations between the cause X and its prediction $X_{\mathcal{H}}$ conditioned on Y and C.I. the corresponding confidence interval, and inference of the corresponding causal direction indicated by an arrow ($\rightarrow$), according to inference rule given in Eq. (8).

Tablica 1 – Binarni usmjereni kauzalni učinci procesnih parametara na VDK vicinalne diketone. Ocijenjeni su Hilbert-Schmidtovi kriteriji neovisnosti (HSIC), parcijalne korelacije $pcor(X_{\mathcal{H}}, Y_{\mathcal{H}})$ i intervali pouzdanosti (C.I.) između uzroka X i posljedice Y u Hilbertovom prostoru, parcijalne korelacije između učinka Y i njegove predikcije $Y_{\mathcal{H}}$ uvjetovane s X i odgovarajućim C.I., $pcor(X, X_{\mathcal{H}} | Y)$ parcijalne korelacije između uzroka X i njegove predikcije $X_{\mathcal{H}}$ uvjetovane s Y i C.I. te zaključak o odgovarajućem kauzalnom smjeru označenom strelicom ($\rightarrow$), prema pravilu zaključivanja danom u jedn. (8).

Effect $Y = \text{VDK}$							
Cause	HSIC	$pcor(X_{\mathcal{H}}, Y_{\mathcal{H}})$	$pcor(Y, Y_{\mathcal{H}} X)$	Confidence interval C.I.	$pcor(X, X_{\mathcal{H}} Y)$	Confidence interval C.I.	Causal direction
T	0.050	0.978	0.779	0.754–0.803	0.681	0.646–0.713	$T \rightarrow \text{VDK}$
X	0.152	0.995	0.997	0.995–0.998	0.987	0.984–0.989	$X \rightarrow \text{VDK}$
S	0.144	0.849	0.889	0.876–0.901	0.872	0.856–0.886	$S \rightarrow \text{VDK}$
E	0.145	0.849	0.890	0.876–0.902	0.872	0.857–0.886	$E \rightarrow \text{VDK}$
CO_2	0.146	0.818	0.890	0.877–0.902	0.910	0.899–0.920	$\text{CO}_2 \leftarrow \text{VDK}$

and are most likely biased. Strong interdependence of the process variables and the quality parameter are evident from partial correlations $pcor(X_{\mathcal{H}}, Y_{\mathcal{H}})$ in RKHS space, given in Table 1.

The coefficients are in the range [0.818–0.995]. The highest RKHS partial correlations are between temperature T and biomass X with VDK quality. Evaluated inference of causal direction between the process variables and quality are based on Kolmogorov's conjecture of complexity metrics due to asymmetric Pearson correlations between conditioned models of cause and effect, $pcor(Y, Y_{\mathcal{H}} | X)$ and $pcor(X, X_{\mathcal{H}} | Y)$, evaluated in RKHS space, as given by Eq. (8). The Pearson correlations and the corresponding confidence intervals are given in Table 1. The inferred directions confirm causal dependence of quality on the process variables, except for erroneously estimated influence of quality on carbon dioxide.

Table 1. Statistical evaluation of causal direction of the process variables on quality $Y = \text{VDK}$ by partial correlation ($pcor$), Hilbert Schmidt independence criteria (HSIC), confidence interval (C.I.), causal direction is indicated by an arrow ($\rightarrow$).

In this example, when process variables are highly correlated, analysis of binary correlations is confounded and causality inferences are unsuitable for process control and/or policy decisions. To de-confound complexity of process variables interactions, binary causal relations are ensembled and balanced yielding a causal Bayes network. Sampled process dynamics, Eq. (13), is discretised and transformed into a data frame, Eq. (14), which complies with the necessary Markovian property. To infer DAG structure, the heuristic algorithm for optimisation of Bayes information criterion BIC was applied.²² The likelihood of the observed data for a potential candidate DAG structure is calculated assuming a Gaussian conditional probability

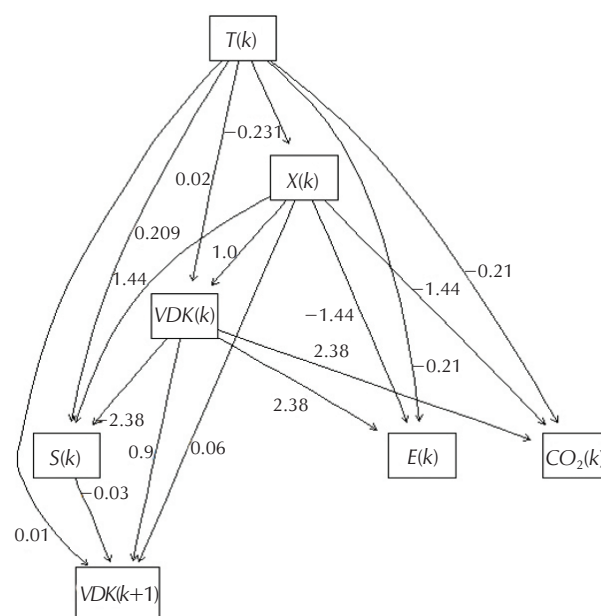


Fig. 3 – Causal Bayes directed acyclic graph (DAG), k is time series sampling index, and the average causal effects are numerically denoted alongside the network edges

Slika 3 – Kauzalni Bayesom usmjereni aciklički graf (DAG), k je indeks uzorkovanja vremenskih serija, a prosječni kauzalni učinci numerički su označeni uz rubove mreže

distribution. The assumed structures are constrained by a priori knowledge of the inferred causal directions (Table 1), except for VDK and CO_2 , and the direct dependence of $\text{VDK}(k+1)$ on $\text{VDK}(k)$ due to ODE discretisation (Eq. (14)). After 195 iterations (tested structures), the inferred most probable DAG based on likelihood is depicted in Fig. 3. It consists of 7 nodes and 16 directed arcs, with an average

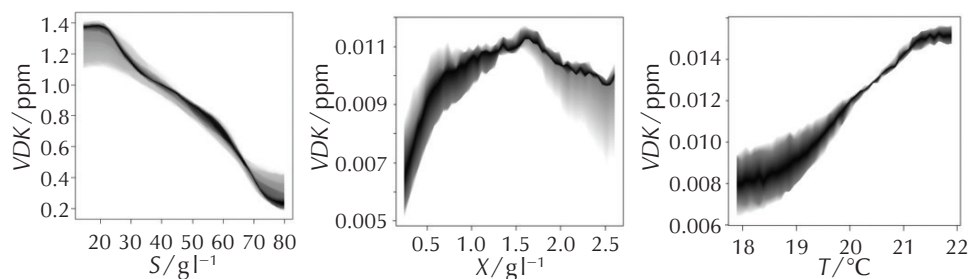


Fig. 4 – Bayesian neural network simulations of causal intervention dependencies of VDK($\text{do}(X=x)$) concentration on the process parameters: S sugar concentration, X biomass concentration, T temperature

Slika 4 – Primjena Bayesove neuronske mreže u simulaciji kauzalne intervencijske ovisnosti koncentracije VDK($\text{do}(X=x)$) o parametrima procesa: koncentracija šećera S , koncentracija biomase X , temperatura T

Markov blanket size of 4.57 and an average branching factor of 2.29. Temperature time transient $T(k)$ is the only manipulative (control input) exogenous variable, which has direct causal effect on all the network nodes, i.e., on each process variable and product quality. The individual direct causal effects (strength and sign of the edges) are estimated by linear ordinary least square (OLS) regression with the standardised data set. Indirect causal path effects are given by product of edge coefficients along the corresponding path. The calculated coefficients can be considered only as a linearised approximation of true non-linear causal dependencies. To assess probabilistic non-linear dependencies, Bayesian neural network (BNN) models were employed.²⁵ BNN model were developed for each DAG causal path with d-separated variables. Due to deconfounded effects, partial dependence plots (PDP) were used to show causal dependencies in the range of each causal factor, as shown in Fig. 4. For example, the causal effect of biomass on product quality changes from positive at lower concentrations to negative at higher biomass concentrations. The “fanned” or “dispersed” graphs represent the probability density functions $P(Y|\text{do}(X), Z)$ of the quality Y on the cause X due to variations in covariates Z .

From the perspective of process optimisation, the sensitivities of the final fermentation states to the initial states were evaluated. Fermentation profiles were simulated by numerically integrating the model's ODE with a 5 % increase in the initial state variables and manipulative process parameters, specifically the heat transfer coefficient k_H and reference temperature T_{ref} . The results, expressed as relative effects on the final states, are presented in Table 2. The most pronounced effect of 5 % perturbations on quality VDK % were observed with changes in the initial temperature and reference temperature. These effects were non-proportional (non-linear): a 5 % increase in each led to a decrease in quality of -13.2 % and -16.4 %, respectively. In contrast, an increase in the heat transfer coefficient produced a positive effect on quality. The non-linear sensitivity results and interference (confounding) of the process parameters justify the need for non-linear data analysis and structural causal modelling (SCM) for process control and optimisation.

Table 1 – Relative sensitivities of the final process states (quality VDK, biomass X , sugar S , ethanol E) caused by 5 % perturbations of the process initial states and process parameters (reference temperature T_{ref} and heat transfer coefficient k_H)

Tablica 1 – Relativne osjetljivosti konačnih stanja procesa (kvaliteta VDK, biomasa X , šećer S , etanol E) uzrokovane 5 %-tnim perturbacijama početnih stanja procesa i procesnih parametara (referentna temperatura T_{ref} i koeficijent prijenosa topline k_H)

5 % perturbations	Perturbed process states			
Initial state	VDK / %	X / %	S / %	E / %
$T(0)$	-13.2	-12.8	-6.7	1.5
$X(0)$	-1.2	-1.3	-0.5	0.1
$S(0)$	5.1	5.6	-3.4	6.9
parameters				
T_{ref}	-16.4	-12.0	-9.1	2.0
k_H	3.6	1.4	22.6	-5.0

Prediction errors of VDK concentrations were compared between a random forest (RF) model based on an “agnostic” approach (using all variables in the prediction set) and the SCM model depicted in Fig. 3. A negligible reduction in error was observed for the SCM-based RF model. For simulated data with 10 % added relative error, the average prediction accuracy of the agnostic RF model was 96 %, while for the SCM-based RF model the error was 94 %. The slightly higher apparent accuracy of the agnostic model is attributed to overfitting bias due to confounding variables, specifically ethanol and CO_2 data.

5 Conclusions

This study presents a modelling and causal analysis of the beer fermentation process, highlighting the crucial role of embedding data in a Reproducing Kernel Hilbert Space (RKHS).

The process data used were generated through multiple simulations of ordinary differential equation (ODE) metabolic based models, with initial condition variations and parameter values derived from laboratory-scale experimental studies reported in the literature. Simulated Gaussian process noise was added to the data to reflect realistic disturbances.

Using the simulation model, sensitivities of the process states were evaluated in response to 5 % perturbations of the initial conditions and key process parameters, including the reference temperature and the heat transfer coefficient. The resulting sensitivity analysis revealed pronounced non-linear relationships between the input perturbations and their effects on the final fermentation state. Among the factors analysed were: temperature, specifically the initial temperature, reference temperature, and heat transfer coefficient, which exert the most significant causal influence on product quality.

These findings were corroborated and further explained by derived Bayesian Causal Network (BCN). The derived Structural Causal Model (SCM) of the beer fermentation process identifies temperature, sugar concentration, and yeast biomass concentration as direct causal factors. These variables are critical for prediction and control, directly impacting beer quality.

The application of this SCM forms a core component of a Digital Twin, enabling real-time monitoring and control of process variables to maintain consistent product quality. Furthermore, the developed causal Digital Twin model for beer fermentation can support innovations in product development, such as the creation of new or market-specific flavours while ensuring high product quality.

A comparison of the random forest prediction errors between the agnostic model and the SCM shows only a negligible decrease in accuracy, likely due to bias from overfitting to non-causal variables. However, the primary value of the SCM lies not in improved prediction, but in its ability to uncover the $P(\text{do}(X=x))$ probability, which enables causal process control modelling and informed decision-making.

More broadly, this approach facilitates the modelling of Bayesian Directed Acyclic Graphs (DAGs), which are fundamental for developing causal AI-based Digital Twins with intervention capabilities in food production and process control systems.

The main reasons are:

- Embedding data into RKHS Hilbert space enables application of linear algorithms for statistical analysis and machine learning (ML) algorithms for pattern discovery through the “kernel trick.”
- The application of the Hilbert-Schmidt Independence

Criterion (HSIC) reveals correlations between non-linear functional dependencies.

- Utilising Kolmogorov’s notion of complexity metrics, the evaluation of the asymmetric properties of conditional distributions $P(X,Y)$ and $P(Y,X)$ allows for the inference of directional orientation between the cause X and effect Y .
- Bayesian Directed Acyclic Graphs (DAGs) facilitate the unconfounded identification and estimation of causal relationships between multiple variables. This is crucial in process control, where understanding unconfounded cause-and-effect relationships is essential for optimising and improving processes.
- Bayesian DAGs provide a probabilistic framework for modelling complex systems. This allows for the incorporation of uncertainty and variability in the process, leading to more robust and reliable control strategies.
- Bayesian methods enable the continuous updating of the Digital Twin model as new data becomes available. This is particularly useful in process control, where conditions can change over time, and the model needs to adapt accordingly.
- Bayesian DAGs support Digital Twin model decision-making under uncertainty by providing a structured way to combine prior knowledge with new evidence. This helps in making informed decisions about process adjustments and interventions.
- Importantly, due to DAGs graph structure, causal Digital Twins offer a clear and intuitive way to visualise the dependencies and relationships between variables. This aids in understanding the process and communicating the control strategy to stakeholders.

The future of Digital Twins in food technology is promising, with emerging trends including enhanced predictive capabilities through advanced instrumental analytics and artificial general intelligence (AGI), new and sustainable food chains under adverse climate changes, for sustainable, safe, improved traceability and transparency in supply chains, and the development of personalised food products based on individual health status. Importantly, there is also an emphasis on reducing food waste and optimising resource use.

From a food company management perspective, the introduction of Digital Twins requires significant investments, company restructuring, and new generation of food engineers educated in AI technologies. However, as technology advances, Digital Twins are expected to become an integral part of the food technology landscape, driving innovation and addressing global food challenges. The potential applications in the food industry are immense and will greatly outweigh the challenges, ultimately increasing quality of life of global human population and environmental protection.

References Literatura

1. W. B. Yooh, H. W. Yang, MDPI Special Issue Applications of Artificial Intelligence in Food Processing and Food Industries, 2024, url: https://www.mdpi.com/journal/processes/special_issues/AI_Food_Processing.
2. Ž. Kurtanek, Causal artificial intelligence models of food quality data, *Food Technol. Biotechnol.* **62** (1) (2024) 102–109, doi: <https://doi.org/10.17113/ftb.62.01.24.8301>.
3. Y. J. Tseng, P. J. Chung, M. Appell, When machine learning and deep learning come to the big data in food chemistry, *ACS Omega* **8** (13) (2023) 15854–15864, doi: <https://doi.org/10.1021/acsomega.2c07722>.
4. D. A. Gee, W. F. Ramirez, A flavour model for beer fermentation. *Journal of the Institute of Brewing* **100** (5) (1994) 321–329, doi: <https://doi.org/10.1002/j.2050-0416.1994.tb00830.x>.
5. C.-M. Agbai, Application of artificial intelligence (AI) in food industry, *GSC Biol. Pharm. Sci.* **13** (1) (2020) 171–178, doi: <https://doi.org/10.30574/gscbps.2020.13.1.0320>.
6. C. Krupitzer, T. Noack, C. Borsum, Digital Food Twins Combining Data Science and Food Science: System Model, Applications, and Challenges, *Processes* **10** (2022) 1781, doi: <https://doi.org/10.3390/pr10091781>.
7. C. Krupitzer, T. Noack, DigiFoodTwin: Digital Biophysical Twins Combined with Machine Learning for Optimizing Food Processing, *Eng. Proc.* **19** (2022) 42, doi: <https://doi.org/10.3390/ECP2022-12623>.
8. M. G. Corradini, A. K. Homez-Jara, C. Chen, Chapter Three - Virtualization and digital twins of the food supply chain for enhanced food safety, *Adv. Food Nutr. Res.* **111** (2024) 71–91, doi: <https://doi.org/10.1016/bs.afnr.2024.06.001>.
9. T. Y. Melesse, C. Franciosi Di Pasquale V., S. Riemma, Analyzing the Implementation of Digital Twins in the Agri-Food Supply Chain, *Logistics* **7** (2023) 33, doi: <https://doi.org/10.3390/logistics7020033>.
10. T. O. Ahmed Monjur, M. Kamruzzaman, Deep learning-based hyperspectral image reconstruction for quality assessment of agro-products, *J. Food Eng.* **382** (2024) 112223, doi: <https://doi.org/10.1016/j.jfoodeng.2024.112223>.
11. N. Harris, V. C. Gonzalez, J. Zhang, A. Pang, C. Hernandez-Brenes, S. Fuentes, Enhancing beer authentication, quality, and control assessment using non-invasive spectroscopy through bottle and machine learning modeling, *J. Food Sci.* **90** (2025) 90:e17670, doi: <https://doi.org/10.1111/1750-3841.17670>.
12. Ž. Kurtanek, Wine quality analysis by structural causal model, *Croat. J. Food Sci. Technol.* **15** (2) (2023) 1–12, doi: <https://doi.org/10.17508/CJFST.2023.15.2.05>.
13. A. Nosedal-Sanchez, C. B. Storlie, T. C. M. Lee, R. Christensen, Reproducing Kernel Hilbert Spaces for Penalized Regression: A Tutorial, *Am. Stat.* **66** (1) (2012) 50–60, doi: <https://doi.org/10.1080/00031305.2012.678196>.
14. S. Pereverzyev, *An Introduction to Intelligence Based on Reproducing Kernel Hilbert Spaces*, Springer, 2022, ISBN 978-3-030-98316-1.
15. J. Pearl, D. Mackenzie, *The Book of Why*, Penguin, UK, 2018, ISBN 9780141982410.
16. J. Suzuki, *Kernel Methods for Machine Learning with Math and R*, Springer, 2022., doi: <https://doi.org/10.1007/978-981-19-0401-1>.
17. J. Mitrovic, D. Sejdinovic, Y. W. Teh, Causal inference via kernel deviance measures, u S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (ur.), *Advances in Neural Information Processing Systems* 31 – 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Neural Information Processing Systems Foundation, Montréal, Canada, 2018, pp. 1–9, url: https://proceedings.neurips.cc/paper_files/paper/2018/file/73fed-7fd472e502d8908794430511f4d-Paper.pdf.
18. J. Pearl, Causal inference in statistics: An overview, *Stat. Surv.* **3** (2009) 96–146, doi: <https://doi.org/10.1214/09-SS057>.
19. J. Johnston, Heineken Marrying Digital Twins with Operational Data to Reduce Energy, *CGT Consumer Goods Technology* **16** (1) (2024).
20. B. Andrés-Toro, J. M. Giron-Sierra, J. A. Lopez-Orozco, C. Fernandez-Conde, J. M. Peinado, F. García-Ochoa, A kinetic model for beer production under industrial operational conditions. *Math. Comput. Simul.* **48** (1) (1998) 65–74, doi: [https://doi.org/10.1016/S0378-4754\(98\)00147-5](https://doi.org/10.1016/S0378-4754(98)00147-5).
21. T. Chen He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho. K. Chen, R. Mitchell, I. Cano, T. Zhou, M. Li, J. Xie, M. Lin, J. Y. Geng, Y. Li, J. Yuan, xgboost: Extreme Gradient Boosting, 2024, doi: <https://doi.org/10.32614/CRAN.package.xgboost>.
22. T. Hayfield, J. S. Racine, Nonparametric Econometrics: The np Package. *J. Stat. Softw.* **27** (5) (2008) 1–32, doi: <https://doi.org/10.18637/jss.v027.i05>.
23. A Karatzoglou, A. Smola, K. Hornik, kernlab: Kernel-Based Machine Learning Lab, 2024, doi: <https://doi.org/10.32614/CRAN.package.kernlab>.
24. K. B. Prakash, *Machine Learning for Industrial Applications*, Wiley-Scrivener, 2024.
25. S. L. Scott, BoomSpikeSlab: MCMC for Spike and Slab Regression, 2023, doi: <https://doi.org/10.32614/CRAN.package.BoomSpikeSlab>.
26. M. Scutari, Learning Bayesian Networks with the bnlearn R Package, *J. Statist. Softw.* **35** (3) (2010) 1–22, doi: <https://doi.org/10.18637/jss.v035.i03>.
27. H. D. Vinod, Updating statistical measures of causal strength, *Sci. Philos.* **8** (1) (2020) 3–20, doi: <http://doi.org/10.23756/sp.v8i1.497>.
28. M. I. Vowels, N. C. Camgoz, R. Browden, D'ya Like DAGs? A Survey on Structure Learning and Causal Discovery, *ACM Comput. Surv.* **55** (4) (2023) 82, doi: <https://doi.org/10.1145/3527154>.
29. J. M. Zamudio Lara, L. Dewasme, H. Hernández Escoto, A. V. Wouwer, Parameter Estimation of Dynamic Beer Fermentation Models, *Foods* **11** (2022) 3602, doi: <https://doi.org/10.3390/foods11223602>.

SAŽETAK

Primjena kauzalne umjetne inteligencije za napredno upravljanje procesa: primjer kakvoće fermentacije piva

Želimir Kurtanjek

Opisane su metodologije temeljene na Hilbertovom prostoru reproduktivne jezgre (RKHS) na primjeru za modeliranje digitalnih blizanaca (DB) i uzročnu analizu procesa fermentacije piva. Model uzima u obzir ključne procesne varijable (temperaturu, biomasu, šećer, etanol i ugljikov dioksid) zajedno s *online* praćenjem kvalitete piva određivanjem koncentracije diketona (VDK). Skup podataka za modeliranje temelji se na integraciji na skupu podataka deset profila fermentacije s poremećenim početnim uvjetima. Fokus je na uklanjanju multikolinearnosti procesnih podataka za razvoj DB modela s intervencijskim potencijalom usmjerenim na optimizaciju kvalitete piva. Rezultati su grafički prikazani kao usmjereni aciklički graf (DAG). Na temelju DAG strukture, Bayesova neuronska mreža (BNN) primijenjena je za grafički prikaz funkcija gustoće vjerojatnosti nelinearnih ovisnosti kvalitete piva o ključnim uzročnim varijablama procesa. Uz interventni potencijal kauzalnih DB modela, naglašena je potencijalna upotreba kauzalnog digitalnog modela blizanaca za protu-činjeničnu analizu sa svrhom poboljšanja upravljanja procesa i moguće inovacije novih proizvoda.

Ključne riječi

BCN Bayesova kauzalna mreža, BIC Bayesova informacijska značajka, RKHS reproducirajući jezgrični Hilbertov prostor, DAG usmjereni aciklički graf, digitalni blizanci, kvaliteta pive

Sveučilište u Zagrebu Prehrambeno-biotehnološki
fakultet, Pierrotijeva 6, 10 000 Zagreb

Izvorni znanstveni rad
Prispjelo 1. travnja 2025.
Prihvaćeno 26. svibnja 2025.