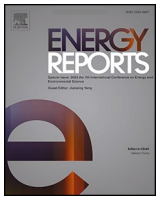
Energy Reports 16 (2026) 109799



Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr



Research Paper



Evaluating regression models and machine learning for long-term energy demand forecasting – Germany

Munzer Osman *, Maria Mercedes Movsessian

School of Business and Economics, RWTH Aachen University, Aachen, Germany

ARTICLE INFO

Keywords:
Energy demand forecasting
Support vector regression
Artificial neural networks
Particle swarm optimization

ABSTRACT

Reliable long-term energy demand forecasting underpins energy-system planning and decarbonization policy design, yet most established methods are calibrated for short horizons and may not capture structural change. This study compares four forecasting models: multiple linear regression (MLR), a vector error correction model (VECM), support vector regression (SVR), and an artificial neural network (ANN). Germany's final energy consumption (FEC) is projected to 2050 from annual data spanning 1992–2021. The SVR and ANN were tuned using particle swarm optimization and Bayesian optimization. Accuracy is assessed using mean absolute error (MAE), mean squared error (MSE), and the coefficient of determination (R^2). The SVR achieves the highest test accuracy (MAE = 9.19, R^2 = 0.974), but its long-term projections substantially exceed the historical range, suggesting extrapolation instability. The ANN attains comparable test accuracy (MAE = 10.05, R^2 = 0.965) with stronger generalization and more plausible projections, and is therefore adopted for the base-case and scenario analyses. MLR is transparent but less accurate, whereas VECM forecasts are sensitive to the assumed cointegration rank in the small-sample setting. The ANN base-case indicates an approximately 12% decline in FEC by 2050 relative to 2008, driven primarily by fuel substitution and efficiency improvements. Across five decarbonization scenarios, projected 2050 FEC ranges from 1476 to 2195 TWh, demonstrating that transition assumptions dominate long-term forecast uncertainty. Even the most ambitious pathway, approximately 39% below 2008, falls short of Germany's 45% FEC reduction target for 2045. The results indicate that generalization and extrapolation behavior are as important as predictive accuracy in model selection for long-term energy transition analysis.

1. Introduction

The global transition toward sustainable energy systems has increased the importance of accurate energy demand forecasting. Reliable long-term projections underpin energy infrastructure investment, energy-security assessment, and decarbonization policy, forming the foundation of effective power system planning. Over a 20- to 30-year forecast horizon, the consequences of forecast error are amplified relative to short-term operational scheduling, since errors propagate into largely irreversible infrastructure commitments. Over such a period, technological innovation (Kluczek et al., 2025), demographic changes (Stermieri et al., 2023), and evolving climate policies are expected to alter energy consumption patterns substantially. The growing complexity of modern energy systems further amplifies this need, as approaches calibrated for short-term stability may fail to capture structural change and nonlinear dynamics (Fodstad et al., 2022).

Energy demand forecasting has been studied using a wide range of techniques, spanning statistical and econometric models, classical machine learning (ML) methods, and more recent deep learning (DL) architectures (Mystakidis et al., 2024). Traditional approaches such as linear regression, cointegration-based models, and time-series techniques remain widely applied in long-horizon settings, due to their interpretability and relatively low data requirements (Ozdemir, 2024). Comparative evidence indicates that such parsimonious models may remain competitive for medium- and long-term forecasting (Mystakidis et al., 2024). ML and DL methods have nonetheless broadened the toolkit, with reported growth in support vector machines (SVM), neural networks (NN), ensemble learners, and recurrent architectures such as long short-term memory (LSTM) networks (Mystakidis et al., 2024).

Support vector regression with exogenous predictors, for example, has achieved a test-set relative error of 1.63% in multivariate monthly electricity sector forecasting, improving on univariate specifications

* Corresponding author.
E-mail addresses: munzer.osman@rwth-aachen.de (M. Osman), movsessian@em-dev.rwth-aachen.de (M.M. Movsessian).

https://doi.org/10.1016/j.egyr.2026.109799
Received 20 January 2026; Received in revised form 4 July 2026; Accepted 17 October 2026
Available online 24 October 2026
2352-4847/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Open access funding enabled and organized by Projekt DEAL.

(Đukanović et al., 2023). LSTM, in turn, has yielded the lowest error among SVM, recurrent, and ensemble methods in short-term load forecasting, with SVM error approximately 13.5% higher (Jain and Gupta, 2024). Most of these advanced methods have been developed for short-term electricity load forecasting (STLF) where high-frequency data permit effective training and frequent recalibration (Jain and Gupta, 2024; Ozerdem et al., 2017). Although they report high short-horizon accuracy, their framing is operational rather than strategic.

The predominance of short-term applications is a material limitation for long-term planning. Reviews note that predictive performance and robustness tend to deteriorate as the forecast horizon lengthens, owing to structural change, non-stationarity, and weakening temporal dependencies (Mystakidis et al., 2024). Strong short-horizon accuracy therefore need not transfer to reliable long-horizon projection. A growing body of work nonetheless shows that ML can perform well over longer horizons under appropriate conditions. National annual demand has been forecast using particle swarm optimized least squares support vector regression (PSO-LSSVR), combined with index decomposition and causality analysis for predictor selection, with improved accuracy under small sample and nonlinear conditions (Yang et al., 2022). A support vector machine optimized by an improved sparrow search algorithm (ISSA-SVM) has reported a mean absolute percentage error (MAPE) of 2.18% for mid- to long-term load forecasting, outperforming several neural and statistical baselines (Li et al., 2022).

Greater model complexity, however, does not guarantee superior long-horizon accuracy. In a comparison of metaheuristic-optimized, neural, and Gaussian-process models for Turkey's annual demand, a stepwise linear regression produced the best forecast (MAPE 2.36%), ahead of the neural and Gaussian-process variants and a genetic-algorithm-optimized model (2.97%) (Ozdemir, 2024).

Hybrid and optimization-based architectures have been proposed to address the nonlinearity of demand series. A model coupling gene expression programming with an adaptive neuro-fuzzy inference system (GEP-ANFIS) has been applied to long-term industrial load forecasting, with the symbolic component informing the fuzzy rule base (Bakare et al., 2024). An evolutionary-strategy LSTM, in which a genetic algorithm tunes the network hyperparameters and principal component analysis reduces feature redundancy, has improved short term load forecasting over least squares SVM and back-propagation baselines (Yingnan and Xiaowei, 2025). Multi-resolution graph neural networks (MRGNN) represent complex multi-scale and spatio-temporal structure in high resolution load data, albeit at intra-day to multi-hour horizons (Mo et al., 2025). These advances confirm the value of hybridization and metaheuristic optimization, yet their demonstrations remain concentrated on short-term, high-frequency settings. The additional design freedom in architectural choices, kernel and regularization parameters, and optimizer settings increases computational burden and amplifies overfitting risk, particularly in the small sample settings typical of long horizon national forecasting (Mystakidis et al., 2024; Jain and Gupta, 2024). Swarm-based optimization, for instance, has been observed to accelerate convergence without consistently surpassing back-propagation accuracy (Ozerdem et al., 2017).

In parallel, long-term energy demand is frequently examined through scenario-based transformation modeling rather than data-driven forecasting. One normative study constructs three greenhouse-gas-neutral pathways for Germany, dominated respectively by direct electrification, hydrogen, and synthetic hydrocarbons (Sensfuß et al., 2021). Direct electrification emerges as the most robust and least cost option, with industrial electricity demand approximately doubling by 2050. Other work couples demand projection with supply-side planning, linking it to cost and carbon-emission objectives through a system-dynamics modeling framework (Guangxiao et al., 2025). Such approaches yield valuable system level insight, but they are not designed to evaluate the empirical forecasting performance or robustness of alternative data-driven models. Comparative evidence on the relative suitability of econometric and optimized ML approaches for long-term demand forecasting therefore remains limited.

Two gaps emerge from this synthesis. First, long-horizon ML applications remain comparatively limited, and the available studies are typically confined to a single framework and national context. China's national demand, for instance, has been addressed in separate single-framework studies (Yang et al., 2022; Li et al., 2022), and Turkey's in a single country model comparison (Ozdemir, 2024). Second, controlled, like-for-like benchmarking of multiple optimized econometric and ML models on a common national dataset, under explicit conditions of structural change and non-stationarity, remains relatively uncommon. Generalization and extrapolation behavior, rather than in-sample fit, are seldom the explicit object of comparison. The present study addresses these gaps through a systematic long horizon comparison of four forecasting models: multiple linear regression (MLR), a vector error correction model (VECM), support vector regression (SVR), and an artificial neural network (ANN). The models are applied to a common German dataset and projected to 2050, with the two ML models tuned using particle swarm optimization (PSO) and Bayesian optimization (BO). Rather than identifying a universally superior forecasting approach, the comparison examines how different model classes balance predictive accuracy, generalization, and extrapolation behavior under conditions of structural change. The resulting projections should therefore be interpreted as conditional on the assumed future trajectories of the explanatory variables rather than as unconditional predictions of Germany's future energy demand. Germany is used as the case study for this analysis.

2. Methods

This study compares two traditional regression approaches, multiple linear regression (MLR) and vector error correction model (VECM), with two widely used machine learning (ML) models, support vector regression (SVR) and artificial neural networks (ANN), for long-term energy demand forecasting in Germany. To enhance predictive performance while controlling model complexity, the hyperparameters of the SVR and ANN models are optimized using particle swarm optimization (PSO) and Bayesian optimization (BO) algorithms. The four models are not intended to be identical in estimation procedure; rather, they represent complementary modeling families evaluated on their ability to explain historical final energy consumption and support conditional long-horizon projections.

2.1. Description of models

2.1.1. Multiple linear regression (MLR)

The MLR examines linear relationships between a dependent variable y and multiple independent predictors x. It is represented as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \quad \varepsilon \tag{1}$$

The coefficients β are estimated using Ordinary Least Squares (OLS) to minimize prediction errors ε. Key assumptions include linearity between predictors and the dependent variable, homoscedasticity of errors, and absence of perfect multicollinearity (Hansen, 2022). In this study, MLR serves as an interpretable benchmark against which the added value of nonlinear modeling approaches can be assessed.

2.1.2. Vector error correction model (VECM)

The VECM, an extension of Vector Autoregression (VAR) models, captures both short-term dynamics and long-term equilibrium relationships among multivariate non-stationary time series (Gao et al., 2025). It is particularly suitable where predictors are interdependent and evolve jointly over time. The VECM is expressed as:

$$\Delta y_t = \alpha\beta' y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \mathrm{det}_t + \varepsilon_t \tag{2}$$

where Δy_t denotes the first difference of the variables; α is the adjust-

ment coefficient representing the speed of convergence toward long-term equilibrium (i.e., the cointegration relationship); β is the cointegration vector; $\beta' y_{t-1}$ is the Error Correction Term (ECT), representing deviations from the long-term equilibrium; Γ_i represents the short-term dynamic coefficients; det_t is the deterministic term including constants, trends, seasonality and structural dummies; p is the lag order; and ε_t is the error term. The application of VECM requires variables to be integrated of order $I(1)$ (Chen, 2024), the presence of cointegration, linear long-run associations among variables, and the absence of major structural breaks (Gao et al., 2025). Given these requirements, the VECM is used primarily to characterize historical long-run equilibrium relationships and to provide an econometric reference case, rather than as a directly symmetric comparator to the conditional-mean models.

2.1.3. Support vector regression (SVR)

The SVR is used to model potentially complex and nonlinear relationships in systems with multiple input features. This study adopts the ε-SVR formulation, which balances model complexity and prediction accuracy through an ε-insensitive loss function. The SVR seeks the flattest hyperplane that contains most of the training data by combining loss minimization with geometrical constraints for effective regression and strong generalization (Wu and Wang, 2023). Nonlinear dependencies are captured using a Radial Basis Function (RBF) kernel, which maps the input variables into a higher-dimensional feature space. The SVR model is governed by three key hyperparameters: the regularization parameter C, the ε-insensitive loss width ε, and the kernel parameter γ.

$$\min \frac{1}{2}||w||^2 + C\sum_{i=1}^{N} \xi_i + \xi_i^* \tag{3}$$

$$K_{\mathrm{RBF}}(x, x') = \quad \exp^{\left(-\gamma\,|x-x'|^2\right)}, \gamma = 1/2\sigma^2 \tag{4}$$

where w is the weight vector, x is the feature vector, σ is the standard deviation, and ξ_i and ξ_i^* are slack variables. The calculated kernel value, which is based on the Euclidean distance $|x - x'|^2$, measures the similarity between two points (x, x′) (Du et al., 2024). The parameter γ, which is inversely proportional to the variance of the kernel σ^2, controls the spread of the kernel (Tharwat, 2019). Higher γ values allow finer discrimination between nearby observations, while lower γ values increase smoothness and reduce model sensitivity, thereby influencing the bias-variance trade-off and generalization performance. Because SVR can fit nonlinear patterns strongly within the observed data range, its long-horizon projections are interpreted with particular attention to extrapolation stability.

2.1.4. Artificial neural network (ANN)

ANNs approximate complex relationships between input features and a target variable through interconnected neurons organized into input, hidden, and output layers (Zivich and Naimi, 2025). Each layer is composed of multiple neurons, each with an associated activation function (Abdullah et al., 2024). Different activation functions are used to map the relationship between the features and the target variable. The neurons between different layers are connected via weights or coefficients (Montesinos López et al., 2022), which define the strength of the connection between these neurons. Additionally, biases are used in ANN to shift the activation function to better fit the modeled data. For a given activation function (φ), the output of the $i^{(th)}$ hidden neuron h_i is computed by applying φ to the weighted sum of its inputs x_j with weights w_{ij} and bias b_i as:

$$h_i = \varphi\left(\sum_j w_{ij} x_j + b_i\right) \tag{5}$$

The contribution of each neuron in the hidden layer is aggregated using weights v_i and bias c to form the output y of the ANN:

$$y = \varphi\left(\sum_i v_i h_i + c\right) \tag{6}$$

To minimize prediction errors, the ANN iteratively updates its weights and biases using backpropagation, which computes gradients of the loss function with respect to each parameter (Montesinos López et al., 2022). In this study, the Adaptive Moment Estimation (Adam) optimizer is employed. The ANN architecture and parameters, such as the activation function type, learning rate, batch size, and number of training cycles (epochs), directly influence its output. During ANN training, the network's learning progress is monitored through the training and validation loss curves. The training loss indicates how well the model fits the training dataset, while the validation loss is used to evaluate the model's ability to generalize to unseen data (Dang Nguyen et al., 2022). This study utilizes TensorFlow to design, train, and test the ANN. Given the limited sample size, the ANN architecture is kept parsimonious, and model selection prioritizes generalization performance rather than in-sample fit alone.

2.2. Optimization methods

Hyperparameter optimization is an integral component of machine learning workflows (Bischl et al., 2023). In this study, PSO and BO are employed to reduce the forecast error, control model complexity, and improve out-of-sample generalization performance. For the ANN model, the hyperparameter search space includes the learning rate and the number of neurons, while for the SVR model the regularization parameter C, the ε-insensitive loss width ε, and the RBF kernel parameter γ are optimized. Hyperparameter selection is based on validation performance to reduce the risk of overfitting and to ensure a consistent comparison between the machine learning models.

2.2.1. Particle swarm optimization (PSO)

PSO is a nature-inspired heuristic algorithm well suited for optimizing nonlinear, multidimensional problems (Twumasi et al., 2024). It initializes a population of candidate particles x_i (representing hyperparameter configurations) within predefined bounds of the solution space. At each iteration t, every particle is evaluated by training the ANN and SVR models on the training dataset and evaluating its performance on the validation dataset using mean absolute error (MAE) as the optimization objective. Each particle retains its personal best position ($p_{i,t}^d$), while the swarm tracks the global best position ($p_{l,t}^d$). Particle velocities $v_{i,t}^d$ and positions $x_{i,t}^d$ are then updated according to inertia ω, cognitive c_1, and social c_2 components (Twumasi et al., 2024):

$$v_{i,t+1}^d = \omega \cdot v_{i,t}^d + c_1 * \quad r_1 \quad * \left(p_{i,t}^d - x_{i,t}^d\right) + c_2 * \quad r_2 \quad * \left(p_{l,t}^d - x_{i,t}^d\right) \tag{7}$$

$$x_{i,t+1}^d = x_{i,t}^d + v_{i,t+1}^d \tag{8}$$

where i is the particle identifier, d represents the number of variables optimized for, and r_1 and r_2 are stochastic components drawn from a uniform distribution over the interval (0, 1). The optimization terminates when no further improvement in the objective function is achieved or when the predefined number of iterations is reached.

2.2.2. Bayesian optimization (BO)

In parallel, BO is employed to optimize the hyperparameters of both the ANN and SVR models using the Optuna framework in Python. BO explores the hyperparameter space by constructing a probabilistic surrogate model to approximate an objective function $f(x_i)$ (Shahriari et al., 2016). An acquisition function uses the surrogate model predictions $\mu(x)$ and uncertainty $\sigma^2(x)$ to balance exploitation and exploration, favoring regions with either low expected error $\mu(x)$ or high uncertainty σ^2.

In this study, BO iteratively evaluates candidate hyperparameter

configurations by training the SVR and ANN models on the training dataset and evaluating their performance on the validation dataset. Validation MAE is used as the objective function throughout the optimization process. After each evaluation, the trained model generates predictions for the training, validation and test datasets to monitor model behavior and generalization performance. A Gaussian Process surrogate model is updated after each evaluation, and the next evaluation point is selected by maximizing the Expected Improvement (EI) acquisition function (Christianson and Gramacy, 2023).

2.3. Interdependence, cointegration, and causality assessment

To investigate the relationships among the variables and evaluate the suitability of the selected predictors for forecasting Germany's final energy consumption (FEC), three complementary analytical tools are applied: correlation analysis, cointegration testing, and Granger causality analysis. Together, these analyses characterize the statistical relationships within the dataset, support the specification of the econometric models, and provide insight into both long-run and short-run dependencies among the variables.

2.3.1. Correlation analysis

A Correlation Matrix (CM) is constructed to explore pairwise linear relationships between the selected variables. The CM is based on Pearson correlation coefficients (r), which range from −1–1, where −1 indicates a perfect negative linear relationship and 1 indicates a perfect positive linear relationship. While correlation provides a first indication of association strength, it does not imply causation or identify long-term equilibrium dynamics. Thus, following (Yang et al., 2022), further tests are conducted to assess predictive and dynamic relationships within the dataset.

2.3.2. Cointegration analysis

Cointegration analysis is used to identify stable long-run equilibrium relationships among non-stationary variables (Wen et al., 2022). While pairwise approaches, such as the Engle-Granger test, assess cointegration between two variables by testing the stationarity of regression residuals (Stock and Watson, 2020), they are limited in multivariate systems. Given the system-level nature of energy demand dynamics, this study applies the Johansen cointegration test, which allows for the simultaneous identification of multiple cointegrating relationships within a multivariate framework (Lyhagen and Baris-Tüzemen, 2025).

The Johansen approach is based on VAR and VECM representations, where long-run information is captured by the impact matrix Πy_{t-1} Johansen (1991) which is equivalent to $\alpha\beta' y_{t-1}$ in the VECM (Eq. 2). The rank of Π determines the number of cointegrating vectors present in the system. The cointegration rank is determined using both the trace statistic and the maximum eigenvalue statistic (Lyhagen and Baris-Tüzemen, 2025). The identified cointegrating relationships are subsequently incorporated into the VECM specification to represent the long-run equilibrium structure of the historical data.

2.3.3. Causality analysis

Granger causality testing is employed to examine the directional predictive relationships among the variables. While cointegration analysis identifies long-run equilibrium relationships, Granger causality focuses on short-run dynamics by assessing whether past values of one variable contain information that improves the prediction of another. Importantly, Granger causality does not establish true causation (Shojaie and Fox, 2022), but rather indicates temporal predictability based on lagged dependencies.

Granger causality essentially compares two regression models:

When only using the lag values of y:

$$y_t = \alpha_0 + \sum_{j=1}^{p} \beta_j y_{t-j} + \epsilon_t \tag{9}$$

And when adding the lag values of other time series x:

$$y_t = \alpha_0 + \sum_{j=1}^{p} \beta_j y_{t-j} + \sum_{j=1}^{p} \gamma_j x_{t-j} + \epsilon_t \tag{10}$$

In this study, Granger causality testing is conducted within the VECM framework, consistent with the cointegration results. The test evaluates whether the inclusion of lagged values of an explanatory variable significantly improves model fit relative to a restricted model, using an F-test. Statistical significance is assessed at the 5% level. The results are used to characterize short-run predictive relationships among the variables and to complement the long-run equilibrium analysis provided by the cointegration tests.

Table 1 summarizes the main characteristics of the two statistical tests.

2.4. Data collection

Energy systems exhibit nonlinear complexity influenced by numerous factors, including economic growth, demographics and technological advancements. This study incorporates nine explanatory features representing macroeconomic conditions, demographic development, energy consumption by carrier, and energy efficiency to model Germany's Final Energy Consumption (FEC): Gross Domestic Product (GDP), Population (POP), Residential Consumption Level (RCL), Oil Products (OP) consumption, Natural Gas (NG) consumption, Electricity (EL) consumption, Energy Efficiency (EE), Biofuels and Waste (BIO) consumption, and Coal and Peat products (CP) consumption. Several explanatory variables represent major components of Germany's final energy consumption. Their inclusion is intentional, as the objective of this study is to compare the forecasting behavior of different modeling approaches under assumed changes in the future energy mix, rather than to establish causal relationships between fully exogenous drivers and aggregate energy demand. A total of 30 annual observations are used, covering the period from 1992 to 2021. All data are obtained from a harmonized statistical source (UNdata).

Machine learning algorithms are sensitive to differences in the scale of numerical inputs, making feature scaling a critical preprocessing step (Pinheiro et al., 2025). Therefore, all the time series are standardized using Z-score normalization prior to model training. Fig. 1 presents the standardized FEC and explanatory variables over the study period. The figure includes both time series plots and inset scatter plots illustrating the relationships between each standardized explanatory variable (horizontal axis) and FEC (vertical axis).

To assess out-of-sample predictive performance and reduce the risk of overfitting, the dataset is divided into training and testing subsets (Joseph and Vakayil, 2022). Specifically, 70% of the observations are used for model training, while the remaining 30% are reserved for out-of-sample testing. During training, the validation set was defined as 20% of the training set. Given the limited sample size, model performance is interpreted primarily in terms of comparative generalization across modeling approaches rather than absolute forecasting accuracy.

Table 1
Main characteristics of granger causality and cointegration tests.

Aspect	Granger Causality	Cointegration Test
Purpose	Short-term predictive power (Granger, 1969)	Long-term equilibrium relationship (Wen et al., 2022)
Data Requirement	Stationary in levels or in differences (Yang et al., 2022)	Non-stationary (Wen et al., 2022), integrated of the same order (Chen, 2024)
Focus	Temporal precedence (lags) (Shojaie and Fox, 2022)	Long-run relationship (Chen, 2024)
Interpretation	x helps predict y (Shojaie and Fox, 2022)	x and y move together in the long run (Chen, 2024)

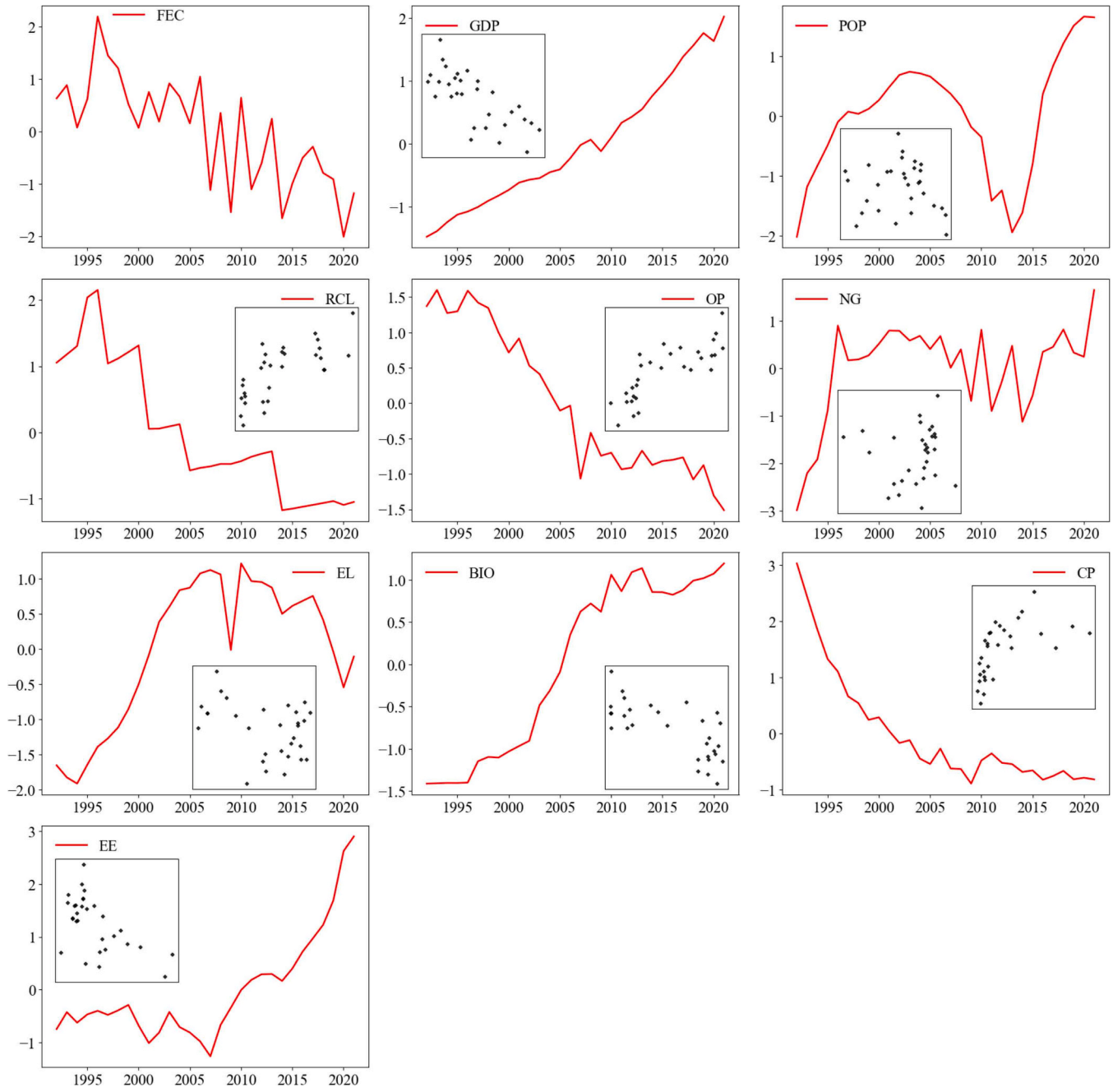


Fig. 1. Historical observations (1992–2021) standardized using Z-score transformation. Insets show the pairwise relationships between each explanatory variable and Germany's final energy consumption. Source: (UNdata).

2.5. *Future trends of the explanatory variables*

To generate long-term forecasts of Germany's FEC, the explanatory variables are extended to 2050 based on official projections, strategic policy documents, and stated assumptions consistent with national energy and climate objectives. The projected trajectories do not represent independent forecasts of the explanatory variables. Rather, they constitute externally specified pathways derived from official projections, policy targets, and transparent assumptions that serve as conditional inputs for the long-term FEC projections. Consequently, the resulting forecasts should be interpreted as conditional on these assumed trajectories.

2.5.1. *Gross domestic product (GDP)*

Germany's GDP growth is projected to recover only gradually, to approximately 1.1% by 2026, owing to ongoing manufacturing challenges, trade uncertainties, and geopolitical tensions (International Monetary Fund IMF, 2025). Despite demographic ageing potentially reducing GDP per capita growth by up to 0.6% annually, average growth rates of 1.2–1.3% per year are expected between 2030 and 2050 (Lizarazo et al., 2022).

2.5.2. *Population (POP)*

Relative to 2023 levels, population is projected to grow by 0.87%, 0.75% and 0.31% by 2030, 2040, and 2050, respectively (Eurostat, 2023). These trends are driven primarily by migration, despite persistently low birth rates and an ageing population structure (Bertelsmann Stiftung Hrsg., 2025).

2.5.3. Oil products (OP) consumption

The amended Federal Climate Change Act (KSG) establishes legally binding, economy-wide greenhouse gas (GHG) reduction targets of at least 65% by 2030 and at least 88% by 2040 relative to 1990, together with net GHG neutrality by 2045 (Federal Republic of Germany, 2019). Meeting these milestones requires a pronounced contraction in fossil-fuel use, of which oil products represent a major share. Accordingly, the reference trajectory assumes OP consumption declines by 50% by 2030, 70% by 2040, and 90% by 2050 relative to 1992, consistent with the decarbonization pathway implied by the KSG.

2.5.4. Coal & peat products (CP) consumption

Coal has long been an essential component of Germany's energy system, although its role has gradually declined. According to the AGEB energy balance (Buttermann, 2024), hard coal and lignite together accounted for about 15% of Germany's primary energy consumption in 2024; coal is used mainly for electricity generation and in iron and steel production. In 2023, coal and lignite accounted for 29.6% of Germany's electricity production (Burger, 2024). However, CP consumption represented only about 2–3% of FEC between 2010 and 2021. Under Germany's Coal Exit Act, coal-fired power generation is to be fully phased out by 2038, with the ambition to achieve this by 2030 (Federal Republic of Germany, 2020). Because the Act governs only coal used for electricity generation, whereas the CP series captures coal across all final energy uses, including residual industrial and metallurgical demand, this study adopts a slightly more conservative assumption, modeling a complete phase-out by 2040, marginally beyond the statutory backstop.

2.5.5. Natural gas (NG) consumption

Over the historical period, natural gas final energy consumption shows no established downward trend, rising from approximately 478 TWh in 1992–661 TWh in 2021 and fluctuating within a 550–660 TWh band over the last two decades (UNdata). The Federal Ministry of Economic Affairs and Climate Action (BMWK) System Development Strategy projects fossil natural gas declining by roughly 25%, 80%, and nearly 100% by 2030, 2040, and 2045 relative to 2021 (Federal Ministry for Economic Affairs and Climate Action BMWK, 2024). Similarly, Öko-Institut reports reductions of 28%, 64%, and 94% by 2030, 2040, and 2050 below 2018 levels (Deloitte, 2023). These figures refer to fossil natural gas across the energy system, including power generation, where the largest reductions occur and which lies outside final energy consumption. In this study, the base case represents a conservative reference trajectory, consistent with the historical persistence of final-energy gas demand, in which NG declines by 4%, 12%, and 20% by 2030, 2040, and 2050 relative to 2021. More ambitious reductions are examined separately through the scenario analysis (Section ↱2.6.), where the accelerated transition pathways reduce final-energy gas demand by up to 90% by 2050.

2.5.6. Electricity (EL) consumption

Germany's Climate Change Act (Federal Ministry for the Environment, 2016) establishes a target for GHG neutrality by 2045, requiring a comprehensive shift from fossil fuels toward renewable electricity, biomass, and green gases. Energy-intensive industries such as steel, cement, and chemicals are expected to electrify core processes, leading to a substantial increase in electricity demand. The BMWK projects electricity consumption to increase by 142% by 2045 relative to 2023 levels (Federal Ministry for Economic Affairs and Climate Action BMWK, 2024). Similarly, the Federal Network Agency (BNetzA) estimates Germany's electricity consumption to reach between 869 and 1200 TWh by 2045 across scenarios (Federal Network Agency Bundesnetzagentur, 2025). In contrast to the more rapid electrification pathways assumed in these projections, this study assumes a more gradual transition, with electricity demand reaching approximately 5% below the lower bound of the BNetzA scenario range by 2050.

2.5.7. Energy efficiency (EE)

Germany's Energy Efficiency Act sets a target to reduce FEC by 45% by 2045 relative to 2008 levels (Federal Republic of Germany, 2023). Continuous improvements in energy efficiency across sectors represent one of the main mechanisms for achieving this target. Accordingly, the base-case trajectory assumes a gradual improvement in the energy-efficiency index of approximately 0.5% per year, reaching about 90% by 2050. The alternative scenarios explore a wider range of efficiency outcomes, with 2050 index values between 85% and 95%.

2.5.8. Biofuels and waste (BIO) consumption

Certified biofuels in Germany demonstrated significant GHG savings in 2023. Compared to a fossil fuel baseline of 94 g CO_2eq/MJ for transport fuels, biofuels achieved actual emissions ranging from near zero to around 35 g CO_2eq/MJ, corresponding to savings of approximately 60–94 g CO_2eq per MJ of fuel replaced (Federal Office for Agriculture and Food BLE, 2024). Reflecting this mitigation potential, this study assumes an expanding role for bioenergy in Germany's energy system, with BIO consumption increasing at an average annual rate of up to 1.5%.

Fig. 2 summarizes the historical observations together with the assumed future trajectories of the explanatory variables through 2050, with all variables scaled for visual comparison.

2.6. Germany's energy demand under alternative scenarios

The trained model is applied to generate conditional projections of Germany's FEC to 2050 under five contrasting decarbonization pathways. Each pathway represents a set of externally specified assumptions regarding future energy-system development rather than an independent forecast of the explanatory variables. Each pathway specifies an internally consistent trajectory for the principal end-use energy carriers, whereas the macroeconomic and demographic predictors, gross domestic product (GDP), population (POP), and residential consumption level (RCL), are held at their base-case values (Section ↱2.5.). This isolates the effect of the energy-system transition on demand, so that differences between scenarios reflect fuel-switching and efficiency assumptions rather than divergent macroeconomic conditions.

Six predictors are treated as scenario levers: oil products (OP), natural gas (NG), electricity (EL), coal and peat products (CP), biofuels and waste (BIO), and energy efficiency (EE). Observed values for 1992–2021 remain unchanged while the extended period 2022–2050 is varied. Each lever is anchored at the last observed year (2021) and interpolated linearly between milestone values defined for 2030, 2040, and 2050. OP reductions are expressed relative to 1992, and NG and EL relative to 2021. BIO follows a constant annual growth rate, CP declines linearly to a full phase-out in the stated year, and EE improves linearly toward its 2050 level. The base-case pathway serves as the reference against which the five scenarios are compared. The scenario assumptions are not intended to predict future policy decisions or technology adoption rates, but to provide a consistent framework for evaluating the sensitivity of the forecasting models to alternative transition pathways. The scenarios are designed to span three principal sources of long-term uncertainty: the pace of decarbonization, the dominant substitution pathway (direct electrification versus low-carbon molecules), and the contribution of demand-side efficiency.

S1 - Accelerated Neutrality is aligned with the 2045 greenhouse-gas (GHG) neutrality target (Federal Republic of Germany, 2019). Fossil carriers decline rapidly, the displaced end-use services are met primarily through electrification, and efficiency gains are strong; FEC consequently falls fastest.

S2 - Delayed Transition assumes slower policy implementation. Natural gas persists as a bridge fuel, coal is phased out slightly beyond the statutory 2038 backstop (Federal Republic of Germany, 2020), and electrification lags, which yields the highest residual demand of the set.

S3 - Deep Electrification maximizes the role of electricity, which

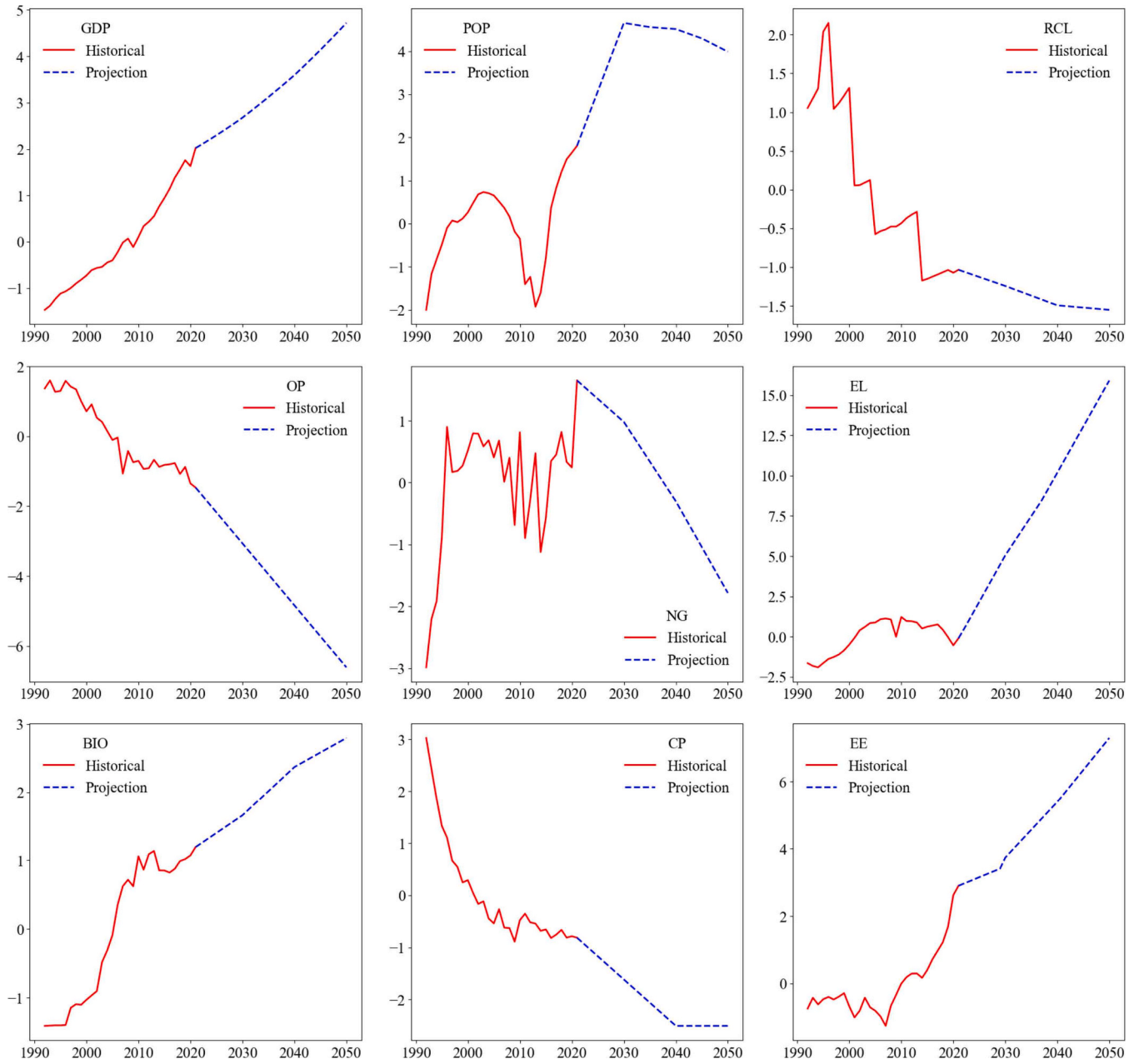


Fig. 2. Historical observations (1992–2021) and assumed future trajectories (2022–2050) of the explanatory variables used to generate the conditional long-term FEC projections. Historical data are obtained from (UNdata), while future trajectories are based on national plans, official projections, and stated assumptions.

substitutes for almost all oil- and gas-based services. Bioenergy remains marginal, and electricity demand approaches the upper range of national projections (Federal Ministry for Economic Affairs and Climate Action BMWK, 2024).

S4 - Green Gas and Bioenergy assigns a larger role to biofuels, biomethane, and green gases where direct electrification is constrained. Electricity grows moderately, and BIO records the highest growth among the scenarios.

S5 - Efficiency / Sufficiency prioritizes demand reduction, in pursuit of the 45% FEC reduction target of the Energy Efficiency Act by 2045 (Federal Republic of Germany, 2023), applying the strongest efficiency and sufficiency assumptions of the set. Electrification is correspondingly restrained, and FEC reaches the lowest level of the set.

Across all pathways the trajectories are coupled to preserve physical consistency. Declining OP and NG reappear, net of efficiency gains, as increased EL or BIO, and the share absorbed by each carrier follows the scenario narrative. This approach represents sector coupling at the aggregate energy-balance level rather than through explicit modeling of individual end-use technologies. For example, heat and transport electrification enters as a reallocation of demand from OP and NG toward EL and BIO, net of efficiency gains, rather than through explicit end-use technologies such as heat pumps or electric vehicles. The pace of decarbonization approximates the speed of this reallocation, with the milestone values in Table 2 setting the rate at which fossil carriers are displaced. These transition rates are imposed exogenously, reflecting plausible policy targets and official projections rather than infrastructural deployment constraints such as grid capacity or charging-network rollout. Consequently, the resulting scenario outputs should be interpreted as conditional on the assumed transition pathways rather than estimates of the achievable rate of systemic reallocation.

The complete set of quantitative assumptions is summarized in Table 2. For each scenario, the projected predictor matrix is

Table 2
Scenario assumptions for the six energy-system levers, 2022–2050.

Scenario	OP (vs 1992)	NG (vs 2021)	EL (vs 2021)	CP phase-out	BIO (%/ yr)	EE (2050)
S1 - Accelerated Neutrality	-60 / −85 / −95	-35 / −70 / −90	+ 40 / + 85 / + 120	2030	+ 2.0	0.93
S2 - Delayed Transition	-25 / −45 / −60	-10 / −20 / −30	+ 15 / + 30 / + 45	2040	+ 1.0	0.85
S3 - Deep Electrification	-50 / −75 / −90	-30 / −60 / −80	+ 45 / + 95 / + 140	2032	+ 0.5	0.90
S4 - Green Gas & Bioenergy	-45 / −65 / −80	-20 / −35 / −50	+ 20 / + 45 / + 65	2035	+ 2.5	0.88
S5 - Efficiency / Sufficiency	-50 / −70 / −85	-25 / −50 / −70	+ 12 / + 22 / + 30	2032	+ 1.0	0.95

Notes: paired values denote the cumulative percentage change at 2030 / 2040 / 2050, with linear interpolation between milestones. OP is measured relative to its 1992 level; NG and EL relative to 2021. CP declines linearly to zero by the stated year and remains zero thereafter. BIO grows at the stated constant annual rate from its 2021 level. EE denotes the efficiency-index level attained in 2050. GDP, POP, and RCL follow the base-case trajectories of Section 2.5 in all scenarios.

standardized and passed to the trained model to obtain FEC. Because several scenario trajectories extend the explanatory variables well beyond their historically observed range, the resulting FEC projections represent conditional scenario outcomes rather than probabilistic forecasts of Germany's future energy demand.

2.7. Robustness and methodological considerations

Long-term energy demand forecasting is subject to uncertainty arising from structural change, policy interventions, and data constraints. In this study, several methodological choices are implemented to enhance robustness and mitigate limitations associated with small sample sizes and long-horizon projections.

First, the analysis is based on annual data for the period 1992–2021, reflecting the availability and consistency of national energy statistics over long time spans. While the resulting number of observations is limited, this temporal resolution is appropriate for long-term demand analysis and avoids inconsistencies that may arise when combining high-frequency data across structurally different regimes. To reduce the risk of overfitting, model complexity is explicitly controlled through regularization in the SVR models, parsimonious network architectures in the ANN models, and systematic hyperparameter optimization based on out-of-sample performance.

Second, the four models belong to different methodological families and are therefore estimated by procedures intrinsic to each framework. This heterogeneity is intentional, as the objective is to compare econometric and machine-learning approaches rather than impose a common estimation algorithm. Comparability is therefore established at the level of the forecasting task and evaluation metrics rather than the estimation procedure itself. All models address the same target variable (FEC), use the same historical dataset (1992–2021), and the same horizon to 2050, and are evaluated using the same metrics (MAE, MSE, and R^2). The conditional-mean models, MLR, SVR, and ANN, estimate FEC from observed explanatory variables and are evaluated using a common 70/30 training-testing split, allowing direct comparison of their out-of-sample performance. The VECM specification differs from the MLR, SVR, and ANN models in two methodological respects. VECM estimation requires variables integrated of the same order. Since the augmented Dickey-Fuller results indicate that NG and CP are stationary in levels, $I(0)$, and POP is $I(2)$ (Section ↱3.1.2.), the VECM is estimated using the seven $I(1)$ series, comprising FEC and six explanatory variables. By contrast, the MLR, SVR, and ANN use all nine explanatory variables. Additionally, the VECM captures the joint short-run dynamics and the long-run cointegrating structure of the time series and is therefore estimated using the full historical sample rather than the training subset. Consequently, the VECM is treated as an econometric reference model, assessed through its in-sample fit, forecast trajectory, and sensitivity to the assumed cointegration rank rather than through direct comparison of held-out errors. This design enables a transparent comparison of fundamentally different modeling approaches while reducing reliance on any single methodology.

Third, the VECM and the ML models respond to structural change through different mechanisms. In the VECM, structural change influences forecasts through the estimated cointegrating relationships and short-run adjustment dynamics, which may no longer hold when externally specified future trajectories are imposed. This behavior is examined through the hybrid VECM configuration presented later in the Results. By contrast, the ANN and SVR receive the structural change through the externally specified explanatory variables and estimate the corresponding level of FEC conditional on those inputs. They therefore do not infer structural breaks from the historical data but instead evaluate the implications of assumed future energy system configurations. Because several explanatory variables represent major components of FEC, part of this relationship reflects the underlying energy balance between energy carriers and aggregate consumption, while the remaining nonlinear component captures interactions learned from the historical data. Consequently, uncertainty is expected to increase primarily when the explanatory variables move beyond the historical range observed during model training.

This difference in extrapolation behavior informs model selection, which is based not only on historical predictive accuracy but also on the plausibility of long-horizon projection. The ANN is preferred because its response to out-of-range inputs remains bounded and consistent with the expected direction of structural change, whereas the SVR amplifies these inputs and produces implausible long-term projections (Section ↱3.4.). This distinction is interpreted as evidence of different extrapolation behavior rather than differences in historical goodness-of-fit alone. Nevertheless, policy-driven changes remain externally imposed through the scenario assumptions and are not endogenously generated by the models.

Fourth, the extension of explanatory variables beyond the historical period is derived from official projections, policy targets, and clearly stated assumptions. These trajectories constitute externally specified scenario inputs rather than forecasts of the explanatory variables themselves. Accordingly, the resulting energy demand projections should be interpreted as conditional on these assumptions rather than as unconditional predictions of Germany's future energy demand.

As an additional robustness check, a hybrid forecasting configuration is tested in which the VECM is augmented with externally projected explanatory variables. The objective is to assess the stability of the estimated cointegrating relationships under policy-driven trajectories that are absent from the historical sample. The hybrid specification is assessed by examining forecast stability and the preservation of long-run equilibrium relationships rather than by predictive accuracy alone.

As the machine-learning models do not provide analytical predictive variances, prediction intervals are constructed using the jackknife + (leave-one-out conformal) procedure (Barber et al., 2021). Each training observation is removed in turn, the model is refitted on the remaining observations, and a leave-one-out residual together with its corresponding prediction are recorded. The interval combines the empirical distribution of leave-one-out predictions with the residual quantiles, providing marginal coverage under exchangeability without distributional assumptions. The same procedure is applied to the SVR and ANN models to preserve comparability. Because conformal validity assumes exchangeability between calibration and prediction data, these intervals formally quantify predictive uncertainty only within the historical predictor range. They therefore capture model and residual uncertainty, but not uncertainty arising from structural change or extrapolation beyond

the observed data.

Finally, the modeling approaches considered in this study exhibit complementary strengths and limitations. Econometric models such as the VECM are informative for analyzing historical equilibrium relationships but are sensitive to structural instability, while nonlinear ML models offer greater flexibility in capturing complex relationships but require careful interpretation when extrapolated beyond the historical data range. The comparative framework adopted in this study is therefore intended to assess the relative performance, robustness, and long-horizon behavior of alternative modeling approaches rather than identify a universally optimal forecasting model.

Fig. 3 summarizes the overall methodological workflow implemented in this study, illustrating the sequence of data preparation, statistical testing, model construction, hyperparameter optimization, and long-term forecasting. The workflow comprises four stages. First, the nine predictors and FEC are assembled and characterized through statistical testing, including stationarity (ADF), correlation, Johansen cointegration, and Granger causality. In parallel, the predictors are extrapolated to 2050 from policy targets, official projections, and stated assumptions. Second, the models are constructed: the data are standardized (Z-score) and split into 70% training and 30% testing subsets, and a hyperparameter search space is defined for the machine-learning models, whereas MLR and VECM are fit directly on the data, by ordinary least squares and Johansen maximum likelihood respectively, and require no hyperparameter tuning. Third, the hyperparameters of the SVR and ANN are optimized: PSO and BO search the predefined space to identify the combinations that minimize the objective function η, which evaluates predictive performance while penalizing the gap between training and validation errors and thereby reduces overfitting risk in the small-sample regime; each configuration is assessed by MAE, MSE, and R^2, and the optimal hyperparameters are used to retrain and validate the final ML models. Fourth, all four fitted models are applied to the extrapolated predictors to forecast FEC to 2050, both as a base case and under the five decarbonization scenarios.

3. Results

This section presents the empirical findings following the methodological workflow described in Fig. 3. It first characterizes the statistical properties of the historical dataset through correlation, cointegration, and Granger causality analyses. The forecasting performance of the econometric and machine-learning models is then evaluated before the selected models are applied to generate conditional long-term projections of Germany's final energy consumption under the assumed transition pathways.

3.1. Relationships and dynamics assessment

This section presents the key findings, focusing on short-term dynamics, long-run equilibrium relationships, and the statistical dependencies between the variables.

3.1.1. Correlation test results

Fig. 4 presents the Pearson correlation matrix (CM) of the study variables. Final Energy Consumption (FEC) exhibits a strong positive correlation with oil products (OP) consumption ($r = 0.81$) and residential consumption level (RCL) ($r = 0.72$), and a strong negative correlation with gross domestic product (GDP) ($r = -0.73$) and biofuels and waste (BIO) consumption ($r = -0.71$). Moderate correlations are observed between FEC and coal and peat products (CP) ($r = 0.56$) and energy efficiency (EE) ($r = -0.58$), while correlations with electricity consumption (EL), population (POP), and natural gas (NG) consumption are comparatively weak.

Several explanatory variables also display strong pairwise linear correlations. GDP is strongly negatively correlated with OP, RCL, and CP, and strongly positively correlated with BIO consumption and EE. OP, RCL, and CP are highly positively correlated with each other and strongly negatively correlated with BIO and EL. These relationships indicate substantial interdependence among several explanatory variables, highlighting the importance of multivariate modeling approaches that account for their joint behavior rather than interpreting pairwise associations in isolation.

3.1.2. Cointegration test results

Prior to conducting the Johansen cointegration analysis, the stationarity properties of the variables are examined using the augmented Dickey-Fuller (ADF) unit root test. The ADF results indicate that all series are non-stationary in levels, except for NG and CP consumption, which are stationary in levels $I(0)$. After first differencing, all variables become stationary except POP, which attains stationarity only after second differencing, $I(2)$. As cointegration analysis requires variables to be integrated of the same order, the series that do not share $I(1)$ status are excluded from the Johansen analysis: NG and CP, which are stationary in levels, $I(0)$, and POP, which is $I(2)$. The analysis therefore proceeds with the seven $I(1)$ series (FEC, GDP, OP, RCL, EL, EE, and BIO).

Lag length selection is critical for VAR-based models, including the Johansen cointegration test, Granger causality testing, and the VECM. Insufficient lags may omit relevant dynamics while excessive lags reduce estimation efficiency (Bruns and Stern, 2019). Although the Bayesian Information Criterion (BIC) suggests an optimal lag length of four, the Johansen trace test exhibits instability at this specification due to the limited sample size (30 observations). Consequently, the cointegration analysis is conducted using two lags.

Table 3 reports the results of the Johansen cointegration trace test, where the null hypothesis H_0 of no cointegration is sequentially tested against alternatives with an increasing number of cointegrating equations (CE). H_0 is rejected when the trace statistic exceeds the 5% critical value. The relative importance of each cointegrating vector is also reported through the corresponding eigenvalues. The results indicate rejection of H_0 for up to four cointegrating equations. At five cointegrating equations, the trace statistic (15.38) falls marginally below the 5% critical value (15.49), and H_0 can no longer be rejected. These results indicate that the seven $I(1)$ variables share a stable long-run equilibrium structure despite exhibiting individual stochastic trends, thereby supporting the use of a VECM framework for analyzing joint long-term dynamics.

3.1.3. Granger causality test results

Cointegration indicates long-run co-movement and can reflect common stochastic trends or underlying behavioral mechanisms, but it does not by itself establish causality (Łasak and Lont, 2020). Granger causality tests are performed to examine directional predictability among the variables while reducing the risk of spurious inference associated with non-stationary time series (Shojaie and Fox, 2022). Given the possibility of delayed effects and feedback mechanisms, the tests are performed for lag orders up to eight. Table 4 reports only cases in which the null hypothesis of no Granger causality is rejected at the 5% significance level.

The results in Table 4 indicate that none of the explanatory variables Granger causes FEC at lag 1. NG consumption Granger causes FEC at lags 2 and 3, while BIO consumption exhibits predictive power at lag 2. RCL and EL consumption Granger-cause FEC only at longer lags (7 and 8), suggesting delayed adjustment effects. No statistically significant Granger causality is detected from GDP, POP, or OP to FEC within the tested lag structure.

These findings indicate that changes in selected energy carriers contain lagged predictive information for FEC, consistent with heterogeneous adjustment speeds across different components of Germany's energy system. At the same time, Granger causality tests focus on linear lagged relationships and may not capture contemporaneous effects or nonlinear interactions among multiple variables (Shojaie and Fox,

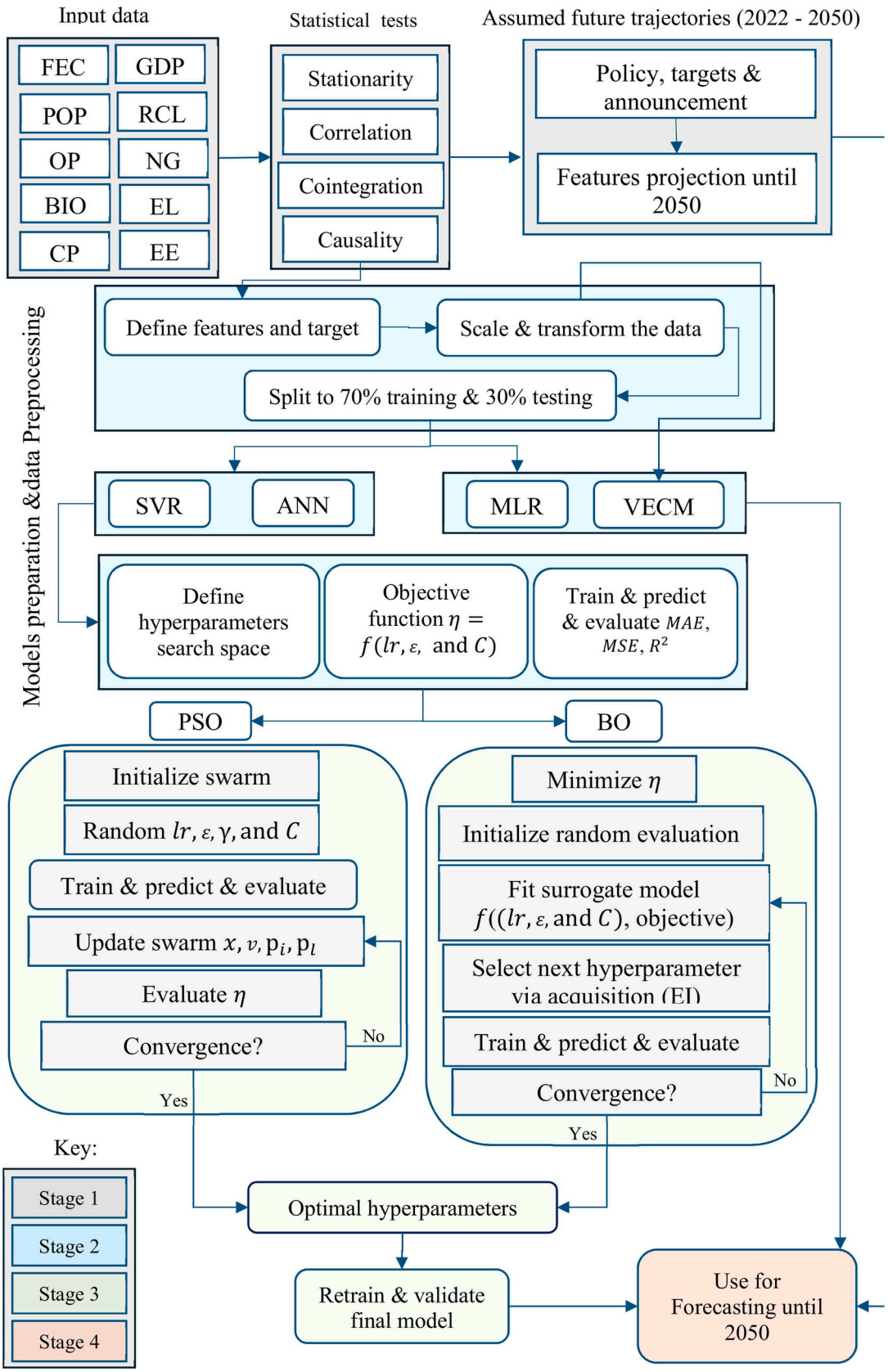


Fig. 3. Methodological workflow adopted in this study. Stage 1: data assembly, statistical characterization (stationarity, correlation, cointegration, and causality), and construction of assumed explanatory variables trajectories to 2050. Stage 2: model setup, including data standardization, training-testing partitioning for the ML models and estimation of the MLR and VECM. Stage 3: hyperparameter optimization of the SVR and ANN using PSO and BO. Stage 4: retraining of the optimized models and generation of conditional FEC projections for the base case and five decarbonization scenarios.

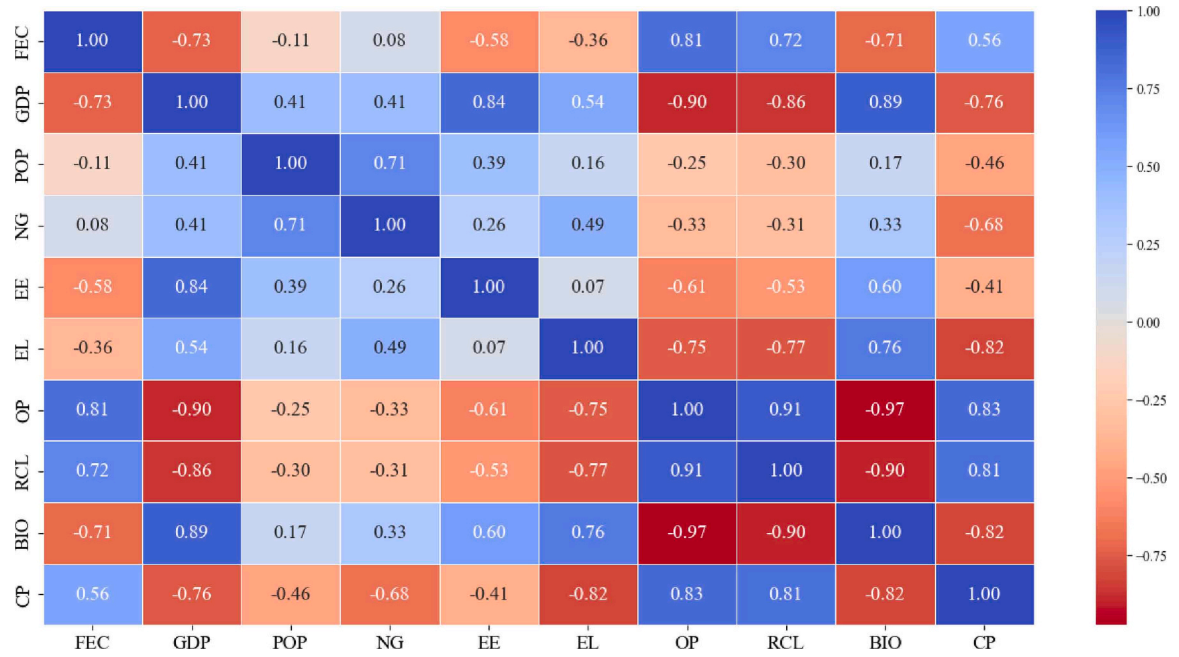


Fig. 4. Pearson correlation matrix (CM) showing pairwise linear relationships among Germany's final energy consumption (FEC), macroeconomic variables, demographic variables, energy carriers and energy efficiency (standardized historical observations, 1992–2021).

Table 3
Results of the Johansen cointegration trace test (2 lags, 5% significance level) for the multivariate system.

No. of CE(s)	Eigenvalue	Trace statistic	0.05 α	Decision
0	0.9986	1561.09	125.61	Reject H_0
1	0.9956	837.15	95.75	Reject H_0
2	0.9544	175.01	69.81	Reject H_0
3	0.8792	91.63	47.85	Reject H_0
4	0.5086	34.56	29.79	Reject H_0
5	0.4308	15.38	15.49	Fail to reject H_0
6	0.0063	0.17	3.84	Fail to reject H_0

Table 4
Summary of granger causality test results for FEC (lags up to 8).

Variable	H_0	Lag	F-stat	P-value
NG	NG does not Granger Cause FEC	2, 3	4.21, 4.49	0.02, 0.01
BIO	BIO does not Granger Cause FEC	2	3.80	0.03
RCL	RCL does not Granger Cause FEC	7	5.71	0.02
EL	EL does not Granger Cause FEC	8	38.12	0.001

2022). Consequently, variables such as GDP, POP, and OP may still improve predictive performance when their combined or nonlinear effects are captured by machine learning models such as SVR and ANN, even though they do not exhibit statistically significant Granger causality.

3.2. Model estimation and evaluation

This section evaluates the performance of the four forecasting approaches using the historical dataset. The MLR, SVR, and ANN are compared on held-out test data, while the VECM is evaluated through its in-sample fit because of its system-based estimation framework. Together, these results provide the basis for selecting the most suitable models for the subsequent long-term projections.

Using historical data from 1992 to 2021, the forecasting performance of MLR, VECM, SVR, and ANN models is evaluated using mean absolute error (MAE), mean squared error (MSE), and the coefficient of determination (R^2). For the SVR and ANN, hyperparameters are optimized prior to performance evaluation.

3.2.1. MLR performance results

The MLR model is evaluated across all possible combinations of the nine predictors, resulting in 511 candidate specifications. Model performance varies substantially across these specifications. Table 5 reports representative high-performing MLR specifications selected to illustrate the trade-off between test MAE, R^2, and model parsimony, highlighting the predictor combinations that provide the highest explanatory and forecasting accuracy.

The full model, including all nine predictors, achieves an R^2 of 0.951 with an MAE of approximately 12.80. Reducing the model specification to five predictors (OP, NG, EL, BIO, and CP) results in only a modest decline in explanatory power ($R^2 = 0.939$) with a comparable MAE. Excluding GDP and population is associated with the largest reductions in model performance, while the removal of RCL and EE leads to relatively minor changes in MAE and R^2. The strong performance of reduced MLR specifications suggests that a relatively small subset of explanatory variables accounts for most of the linear variation in FEC within the historical sample.

Estimated 95% confidence intervals for the standardized MLR coefficients (in TWh per standard deviation) confirm that OP [110.7, 182.6], NG [23.0, 56.3], EL [15.7, 65.3], BIO [74.4, 131.1], and CP [33.7, 69.0] are each significant at the 5% level, whereas GDP, POP, RCL, and EE are not. The limited statistical significance of the macroeconomic and demographic predictors is consistent with the Granger-causality results (Section ↱3.1.3.). The strong contribution of the

Table 5
Forecasting performance of the best-performing and most parsimonious MLR specifications.

Features	MAE	R^2
GDP, OP, NG, EL, EE, POP, RCL, BIO, CP	12.795	0.9507
GDP, OP, NG, EL, EE, POP, BIO, CP	12.821	0.9502
GDP, OP, NG, EL, POP, BIO, CP	12.074	0.9496
OP, NG, EL, POP, BIO, CP	13.523	0.9434
OP, NG, EL, BIO, CP	13.132	0.9388

energy-carrier variables should, however, be interpreted in the context of the modeling framework, as these variables also represent major components of Germany's FEC rather than fully independent explanatory drivers.

3.2.2. VECM performance results

Based on the presence of five cointegrating relationships identified in Section ↱3.1., a VECM with two lagged difference terms is estimated to capture long-run equilibrium relationships and short-run dynamics among the $I(1)$ variables. Table 6 reports the estimated cointegration vectors (β coefficients), where each cointegration equation (CE) is normalized by setting one variable's coefficient to unity for identification.

The estimated cointegration vectors indicate multiple long-run equilibrium relationships among the variables. In the FEC-normalized relationship, BIO and EE enter with positive coefficients (0.852 and 0.193), indicating that these variables co-move with aggregate energy consumption in the long run. In other cointegrating relationships, BIO exhibits positive association with OP and negative association with EL, illustrating the heterogeneous equilibrium relationships identified by the Johansen procedure. These coefficients represent statistical equilibrium relationships rather than causal effects and should be interpreted accordingly.

Incorporating the short-term dynamics and fitting the VECM to the historical data yield the results shown in Fig. 5 below.

On the full estimation sample, the VECM attains an in-sample MAE of 21.325 and an R^2 of 0.878, indicating that it reproduces the historical series moderately well while capturing the long-run equilibrium structure. Because the VECM is estimated on the complete sample and models the joint dynamics of the multivariate system rather than conditioning on observed explanatory variables, this in-sample fit is not directly comparable to the held-out test performance of the MLR, SVR, and ANN models. The VECM is therefore treated as an econometric benchmark for analyzing long-run equilibrium dynamics rather than as a directly competing predictive model (Section ↱2.7.). The moderate in-sample fit is consistent with the limited sample size and the relatively high parameterization of the VECM, as such models typically require longer time series to capture adjustment dynamics reliably (Canepa, 2022).

3.2.3. SVR performance results

Table 7 summarizes the best-performing linear-kernel SVR configurations obtained using PSO and BO optimization. Across both optimization methods, test-set performance is highly consistent, with R^2 values clustered around 0.95. PSO yields the strongest test performance, achieving a minimum MAE of 11.32 and a maximum R^2 of 0.9535, marginally outperforming BO, which achieves slightly lower validation errors but marginally reduced test accuracy. The convergence of the two optimization strategies toward similar hyperparameter regions indicates that the linear SVR solution is robust with respect to the optimization procedure.

In contrast, the SVR model with an RBF kernel exhibits further performance improvements, achieving a test MAE of 9.19 and an R^2 of 0.9735. This result highlights the importance of capturing nonlinear relationships in modeling final energy consumption, although the greater flexibility of the nonlinear kernel also motivates careful evaluation of its extrapolation behavior in the long-term forecasting analysis.

Table 6
VECM cointegration matrix (β coefficients, Z-standardized).

Variable	CE.1	CE.2	CE.3	CE.4	CE.5
FEC	1.00				
GDP		1.00			
OP			1.00		
RCL				1.00	
EL					1.00
EE	0.193	-0.431	0.046	-0.044	0.236
BIO	0.852	-0.661	1.061	1.097	-1.452

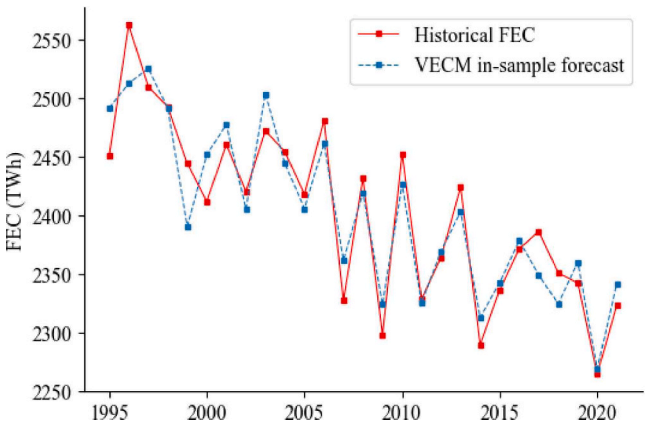


Fig. 5. Comparison of observed and VECM in-sample estimates of Germany's final energy consumption (FEC) (1995–2021). The comparison starts in 1995 because the VECM specification with two lagged differences requires the 1992–1994 observations to initialize the lag structure.

3.2.4. ANN performance results

The baseline ANN with a single hidden layer achieves a test MAE of 14.34 and an R^2 of 0.941, performing below the best MLR and SVR specifications. When trained on the full training dataset, ANN performance improves to an MAE of 12.76 and an R^2 of 0.951, comparable to the best MLR and linear SVR models. Further hyperparameter optimization using BO yields a substantial performance gain, with an MAE of 10.05 and an R^2 of 0.965, making the ANN competitive with the SVR-RBF model.

Increasing network depth or introducing more complex activation functions does not yield additional accuracy gains, indicating that the dominant relationships in the data are adequately captured by relatively simple network structures. Together with its strong held-out performance, this suggests that a parsimonious ANN provides a favorable balance between predictive accuracy and generalization for the present small-sample forecasting problem. This finding is consistent with the strong performance of linear and kernel-based models in small-sample, long-horizon settings.

3.3. Comparative analysis of model performance

Following the individual evaluation of each forecasting approach, this section compares the predictive performance of the competing models. The objective is not only to identify the most accurate model within the historical sample, but also to assess how model flexibility influences generalization performance and, ultimately, the suitability of each approach for long-term forecasting.

The following comparison is restricted to the MLR, SVR, and ANN models, each of which predicts FEC from observed predictor values and is scored on the common held-out test set. The VECM is excluded from this ranking because it is estimated on the full sample and forecasts the joint system dynamics rather than conditioning on observed predictors. Hence, a shared test-set error would not constitute a like-for-like comparison. Instead, the VECM is assessed separately through in-sample fit (Section ↱3.2.2.), forecast behavior, and sensitivity to the assumed cointegration rank (Fig. 8 and Fig. 9).

Fig. 6 and Fig. 7 compare the predictive accuracy of the best-performing MLR, SVR, and ANN models using MAE for the training and testing datasets, respectively. On the training set (Fig. 6), the SVR with an RBF kernel achieves the lowest MAE_{train} (1.78), followed by the MLR model (MAE_{train} = 5.23), while the ANN records the highest training error (MAE_{train} = 8.59). The substantially lower training

Table 7
Performance of linear-kernel SVR model with PSO and BO hyperparameter optimization across training, validation, and testing datasets.

iter	C	ε	MAE_{train}	MSE_{train}	$R^2{}_{train}$	MAE_{val}	MSE_{val}	$R^2{}_{val}$	MAE_{test}	MSE_{test}	$R^2{}_{test}$
PSO											
354	400.46	0.0394	3.8439	23.0572	0.99484	11.92346	221.3926	0.9331	11.50915	188.7447	0.95349
323	931.62	0.04091	3.90158	22.43423	0.99498	11.68545	214.1778	0.93528	11.38958	188.7679	0.95349
554	**484.07**	**0.04177**	**3.93873**	**22.15672**	**0.99504**	**11.53349**	**209.9799**	**0.93655**	**11.32137**	**188.8142**	**0.95348**
291	322.35	0.04178	3.93692	22.16652	0.99504	11.54311	210.3022	0.93645	11.33039	188.9269	0.95345
624	56.65	0.04137	3.91753	22.29598	0.99501	11.61909	212.4097	0.93582	11.36475	189.0249	0.95342
BO											
44	**1000**	**0.04804**	**4.18021**	**21.30984**	**0.99523**	**10.55896**	**185.3028**	**0.94401**	**11.28001**	**195.3806**	**0.95186**
37	541.41	0.05144	4.25439	21.31366	0.99523	10.33724	178.6155	0.94603	11.524	203.7357	0.9498
39	77.88	0.00027	3.12914	36.08782	0.99193	13.78172	320.1944	0.90325	12.45187	210.3335	0.94817
25	69.86	0.05611	4.34877	21.80642	0.99512	10.25509	171.0595	0.94831	12.02216	213.8667	0.9473
46	33.54	0.05428	4.33906	21.90226	0.9951	10.09253	168.6039	0.94905	11.86307	214.8849	0.94705

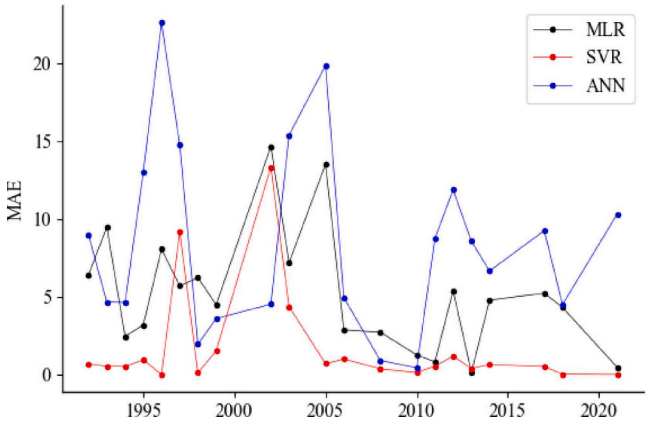


Fig. 6. Training MAE comparison of the best-performing MLR, SVR, and ANN models.

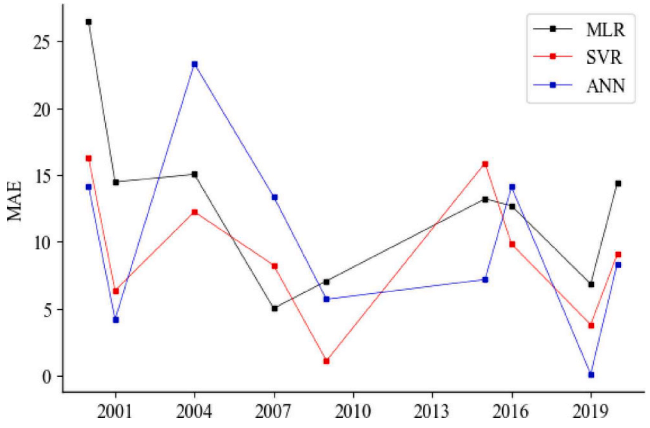


Fig. 7. Testing MAE comparison of the best-performing MLR, SVR, and ANN models.

error of the SVR reflects its greater flexibility in capturing nonlinear relationships within the historical sample. Conversely, the ANN exhibits a more conservative fit, suggesting stronger regularization and a lower tendency to overfit.

On the testing set (Fig. 7), SVR-RBF maintains the highest predictive accuracy with an MAE_{test} of 9.19. The ANN improves its relative ranking in the out-of-sample evaluation, achieving an MAE_{test} of 10.05 and outperforming the MLR model (MAE = 12.80). Although the SVR remains the most accurate predictor on the held-out observations, the considerably smaller difference between the SVR and ANN on the testing set than on the training set indicates that both nonlinear models generalize well despite their different fitting behavior.

Overall, the comparative results indicate that nonlinear modeling approaches provide clear advantages in modeling Germany's long-term energy demand. While the SVR-RBF achieves the highest predictive accuracy across both training and testing datasets, the ANN combines competitive predictive performance with a more conservative fitting behavior that subsequently translates into more stable long-term extrapolation. In contrast, the MLR provides greater transparency and interpretability but exhibits lower predictive accuracy than the nonlinear approaches.

3.4. Forecasting Germany's long-term final energy consumption (FEC)

Having established the predictive performance of the competing models, this section examines their long-term forecasting behavior. Unlike the previous evaluation, which focuses on historical predictive accuracy, the emphasis here is on the plausibility and robustness of projections under future structural change.

3.4.1. VECM forecasting results

The VECM is first used to generate out-of-sample forecasts of Germany's FEC for the period 2022–2050. Fig. 8 presents the forecast trajectory obtained under the baseline specification with cointegrating relationships. Under this configuration, the model projects an increasing trend in FEC over the forecast horizon.

To assess the sensitivity of the VECM forecasts to the assumed long-run equilibrium structure, the cointegration rank is reduced from five to two. As shown in

Fig. 9, this alternative specification produces a fundamentally different projection. In addition to increasing in-sample MAE by approximately 25%, the model now projects a continuous decline in FEC, reaching approximately 750 TWh by 2050.

The marked divergence between Figs. 8 and 9 demonstrates that the

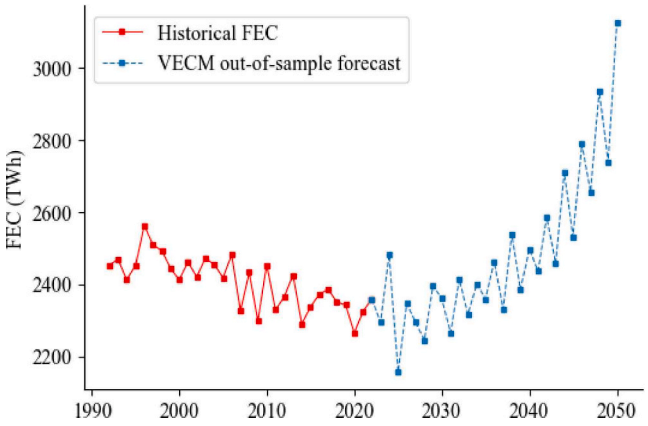


Fig. 8. Out-of-sample VECM forecast of Germany's final energy consumption (FEC) between 2022 and 2050 assuming five cointegrating equations.

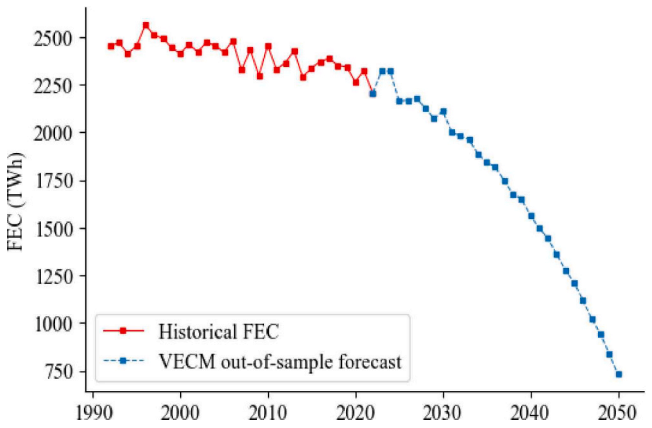


Fig. 9. Out-of-sample VECM forecast of Germany's final energy consumption (FEC) between 2022 and 2050 assuming two cointegrating equations.

VECM forecasts are highly sensitive to the specification of the long-run equilibrium structure. Given that the two models differ only in the assumed cointegration rank, this result illustrates how uncertainty in the estimation of long-run relationships propagates into substantially different long-term forecasts.

A hybrid forecasting approach is additionally tested by augmenting the VECM with externally projected variables as exogenous inputs. However, the resulting long-term forecasts prove unstable. The externally projected exogenous series no longer preserve the historical cointegrating relationships, leading to a breakdown of the estimated long-run equilibrium structure. As a result, the VECM is unable to accommodate policy-driven structural changes embedded in the future projections. This outcome is consistent with the documented sensitivity of VECM frameworks to structural instability, particularly when long-term forecasts extend beyond the historical regime used for model estimation (Gao et al., 2025).

MLR, SVR and ANN long-term forecasting results

Fig. 10 compares long-term forecasts of Germany's FEC generated by the MLR, SVR, and ANN models for the period 2022–2050.

By 2050, the MLR model estimates FEC at approximately 2130 TWh. The estimated 95% prediction interval achieves an empirical coverage of 0.89 (8 out of 9 held-out observations). Although the limited size of the test dataset precludes a rigorous assessment of calibration, the resulting interval remains consistent with the expected uncertainty associated with the model's historical residual variance.

In contrast, the SVR model forecasts a sharp increase in FEC, reaching approximately 3760 TWh by 2050. Applying the jackknife + procedure (Barber et al., 2021), all test-set observations fall

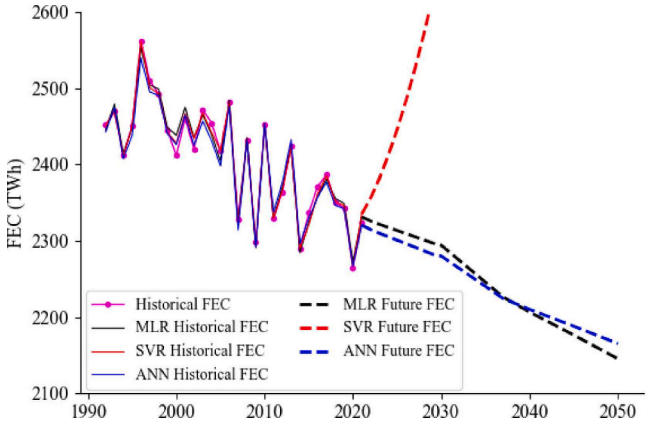


Fig. 10. Long-term forecasts of Germany's final energy consumption (FEC) generated by the MLR, SVR, and ANN models (2022–2050).

within the 95% prediction interval; with only nine test points this is consistent with the procedure's coverage guarantee but does not by itself establish calibration. The SVR prediction interval ranges from 2716 to 3829 TWh. Although all test observations are covered by the interval, the pronounced asymmetry and exceptionally large width indicate substantial disagreement among leave-one-out model refits once the predictor values move beyond the historical range. This interval therefore primarily reflects extrapolation instability rather than credible uncertainty surrounding future demand.

The ANN projects FEC of approximately 2140 TWh by 2050, corresponding to a reduction of around 12% relative to 2008 levels (2432 TWh). The 95% prediction interval is approximately 1909–2505 TWh in 2050. Compared with the SVR, the ANN interval is substantially narrower while remaining within a physically plausible range of future demand, consistent with the stronger generalization performance observed in the historical evaluation. Notably, the entire interval remains above Germany's target of reducing FEC by 45% by 2045.

Both the MLR and ANN models project a gradual decline in FEC over the forecast horizon. By contrast, the SVR predicts a substantial increase in demand. These contrasting projections arise despite the models exhibiting similar historical predictive accuracy, highlighting that strong in-sample and out-of-sample performance alone does not guarantee robust long-term extrapolation. Whereas the MLR and ANN respond to the projected reductions in fossil-fuel consumption and improvements in energy efficiency with declining demand, the nonlinear RBF kernel amplifies the projected increases in EL and BIO consumption, producing an implausible growth trajectory. This divergence reinforces the importance of evaluating forecasting models not only on predictive accuracy but also on the plausibility of their extrapolation behavior.

Permutation importance (Fumagalli et al., 2023) for the ANN, computed as the increase in test-set MAE when each standardized predictor is randomly permuted, identifies OP, BIO, EL, NG, and CP as the most influential predictors of FEC, with effects ranging from a ΔMAE of approximately 114 TWh for OP to approximately 12 TWh for CP. By contrast, GDP, POP, RCL, and EE show limited importance (ΔMAE approximately 9 TWh or less), suggesting that the historical variation in FEC is explained primarily by changes in the composition of energy carriers rather than by macroeconomic or demographic variables.

3.5. Germany's energy demand under alternative scenarios

Because future energy demand depends strongly on the pace and direction of the energy transition, the ANN model is finally applied under the five alternative transition pathways described in Section ↱2.6. The objective is to evaluate the sensitivity of long-term demand projections to different decarbonization assumptions rather than to identify a single deterministic forecast.

Fig. 11 presents the historical final energy consumption (FEC) trend from 1992 to 2021 and the projected trajectories to 2050 under the base case and five transition scenarios. The scenarios indicate divergent decarbonization pathways, with the strongest FEC reduction observed under the efficiency-sufficiency scenario.

The scenario results in Fig. 11 reveal substantial divergence in projected FEC after 2021. Under the base case, ANN-predicted FEC declines from 2323 TWh in 2021 to approximately 2140 TWh in 2050, corresponding to a reduction of 8% relative to 2021 and 12% relative to 2008. The delayed transition scenario produces the highest long-term demand (2195 TWh), whereas the deep electrification and green-gas scenarios reduce FEC to approximately 1999 and 2085 TWh, respectively.

The accelerated neutrality scenario initially produces the strongest decline, reaching 1871 TWh around 2040, before increasing slightly thereafter to 1941 TWh by 2050. This non-monotonic behavior reflects the increasing contribution of electricity and bioenergy to final energy demand, which partially offsets reductions achieved through fossil fuel substitution during the later stages of the transition. The efficiency-and-

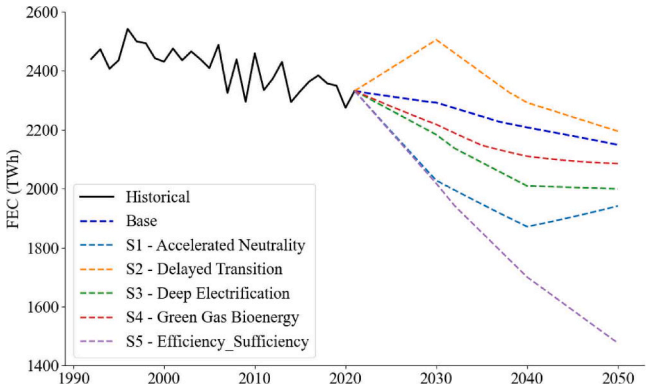


Fig. 11. Scenario-based uncertainty analysis of Germany's final energy consumption (FEC) using the trained ANN model. Historical data are fixed for 1992–2021, while projections for 2022–2050 reflect alternative transition pathways: base case, accelerated neutrality, delayed transition, deep electrification, green gas and bioenergy expansion, and efficiency and sufficiency. Projected 2050 FEC ranges from 1476 to 2195 TWh, highlighting the sensitivity of long-term demand forecasts to transition assumptions.

sufficiency scenario produces the lowest long-term demand, reducing FEC to approximately 1476 TWh by 2050. This corresponds to reductions of approximately 36.5% relative to 2021 and 39.3% relative to 2008.

Overall, the 2050 ANN projections range from 1476 to 2195 TWh across scenarios, producing an uncertainty range of 719 TWh. The magnitude of this spread exceeds the differences between the competing forecasting models themselves, indicating that long-term projections are influenced more strongly by assumptions regarding future transition pathways than by the choice of forecasting algorithm. The results therefore suggest that long-term energy demand forecasts should be interpreted as conditional on explicit transition assumptions rather than as unique predictions. Among the scenarios considered, only the efficiency-and-sufficiency pathway approaches Germany's long-term energy efficiency objectives, although even this scenario remains below the targeted 45% reduction relative to 2008 by 2045. Table 8 summarizes the projected reductions in final energy consumption (FEC) under the different forecast scenarios.

4. Discussion

The comparative analysis conducted in this study highlights important methodological insights, structural dynamics, and practical implications for long-term energy demand forecasting in the context of Germany's energy transition. The results collectively indicate that Germany's final energy consumption (FEC) is weakly responsive to short-run macroeconomic fluctuations and is instead predominantly shaped by long-run structural changes within the energy system.

The presence of multiple cointegrating relationships suggests that long-run equilibrium adjustment occurs primarily through substitution among energy carriers rather than through aggregate expansion of energy demand. In particular, the observed relationships between electricity (EL), coal and peat products (CP), and oil products (OP) point to gradual reallocation processes within the energy mix. Granger causality analysis further reveals heterogeneous adjustment speeds across energy carriers. Natural gas (NG) influences FEC at relatively short horizons (2–3 lags), consistent with its role as a flexible energy carrier responsive to supply and price shocks, whereas EL exhibits delayed responses (around 8 lags), reflecting stronger infrastructural, regulatory, and investment rigidities. Together, these dynamics suggest that Germany's energy transition is characterized by gradual adjustment and delayed system responses, reinforcing the importance of long-term planning horizons in energy policy design.

Table 8
Summary of forecast final energy consumption (FEC) reductions across the scenario-based uncertainty analysis.

Scenario	FEC in 2050 [TWh]	Change vs 2021	Change vs 2008
Base	2140	-8.0%	-12.0%
S1 - Accelerated Neutrality	1941	-16.5%	-20.2%
S2 - Delayed Transition	2195	-5.5%	-9.7%
S3 - Deep Electrification	1999	-14.0%	-17.8%
S4 - Green Gas and Bioenergy	2085	-10.3%	-14.3%
S5 - Efficiency and Sufficiency	1476	-36.5%	-39.3%

The VECM results underscore important methodological limitations for long-horizon forecasting in rapidly transforming energy systems. The strong sensitivity of forecasts to the assumed cointegration rank highlights the challenge of estimating stable long-run equilibrium relationships using limited historical data. In small-sample settings (30 observations), Johansen-based cointegration tests are known to exhibit low statistical power, increasing uncertainty in rank determination and long-run inference (Canepa, 2022). This issue is evident in the markedly different long-run trajectories obtained under alternative cointegration specifications (Fig. 8 and Fig. 9). Moreover, VECM forecasting relies on the assumption that historically estimated equilibrium relationships remain stable over time. Given the scale and pace of Germany's policy-driven energy transition, this assumption becomes increasingly restrictive. While VECM remains valuable for analyzing historical long-run relationships and adjustment mechanisms, its application to long-term forecasting under structural change should therefore be treated with caution.

In contrast, the machine learning models demonstrate greater flexibility in accommodating nonlinear relationships but also exhibit distinct limitations. The SVR model achieves strong in-sample and test performance. However, its long-term projections are generated under conditions that deviate substantially from the historical data range. The sharp increase in Germany's FEC projected by the SVR reflects its sensitivity to nonlinear interactions among rapidly expanding energy carriers, particularly EL and BIO. While this behavior highlights the SVR's capacity to capture complex dependencies, it also illustrates the risk of amplifying growth patterns when extrapolating far beyond historically observed patterns (Yu et al., 2024). As such, SVR appears well suited for short- to medium-term forecasting, whereas its long-horizon projections require careful interpretation within a structural and policy context.

The ANN model shows a more balanced performance profile, combining strong out-of-sample accuracy with a long-term forecast trajectory that remains consistent with observed structural trends. Although its training accuracy is lower than that of the SVR, the ANN demonstrates stronger generalization while maintaining physically plausible long-term projections, making it the most reliable model for long-term energy demand forecasting within this study. Importantly, the results show that a parsimonious ANN architecture optimized through Bayesian optimization is sufficient to capture the dominant relationships in the data. This is in line with literature that indicates additional model complexity does not necessarily yield further performance gains in small sample and long-horizon settings (Ozdemir, 2024).

The five leading predictors from the ANN model coincide with the energy-balance components that jointly account for approximately 94% of historical FEC. However, their dominance partly reflects this accounting structure rather than independent causal influence. Consequently, the permutation importance results should be interpreted as measures of predictive contribution within the fitted ANN rather than as evidence of causal importance. The limited importance of GDP, POP, and RCL is nonetheless consistent with Germany's structural shift toward a service-oriented economy and suggests that FEC is increasingly decoupled from conventional macroeconomic indicators. The low

standalone importance of EE does not imply that efficiency is immaterial to future demand. Over the historical period, EE is highly correlated with GDP and BIO ($r = 0.84$ and 0.60), meaning that part of its predictive contribution is absorbed by these variables during model training. In the scenario analysis, EE varies independently and materially affects projected FEC.

From a policy perspective, the conditional long-term projections generated in this study indicate that Germany is unlikely to achieve its stated FEC reduction targets under current trajectories. The base-case projected reduction of approximately 12% by 2050 relative to 2008 levels (to about 2140 TWh) falls well short of the government's target of a 45% reduction by 2045 (Federal Republic of Germany, 2023). Given that historical reductions had reached only around 4% by 2021, achieving the stated policy targets would therefore require substantially stronger and more coordinated interventions. These findings underscore the need for accelerated fossil fuel phase-outs, systematic improvements in energy efficiency, and sustained policy support for low-carbon energy carriers. At the same time, the increasing role of EL and BIO highlights the importance of managing system integration challenges, including grid stability, infrastructure adequacy, and industrial competitiveness.

The scenario analysis sharpens this conclusion. Across the five pathways, projected 2050 FEC spans 1476–2195 TWh, a range of 719 TWh. This spread demonstrates that uncertainty associated with future transition assumptions exceeds the differences among the plausible model projections (MLR, ANN), highlighting the importance of scenario uncertainty in long-term forecasting. The base-case forecast should therefore not be interpreted as a deterministic prediction, but rather as one pathway within a broader range of plausible trajectories. Critically, even the most ambitious pathway, efficiency and sufficiency (S5), achieves a reduction of only 39.3% relative to 2008, short of the 45% target, hence, none of the tested combinations of fuel substitution and efficiency closes the gap by 2045. The accelerated-neutrality pathway (S1) is instructive: FEC declines until approximately 2040 but rises thereafter, as growing electricity and bioenergy demand partially offsets the contraction in fossil carriers. This behavior suggests that electrification-led substitution alone has limited capacity to lower aggregate demand, and that meeting the national target would require additional demand-side and sufficiency measures beyond fuel-switching. The delayed-transition pathway (S2), with the highest residual demand among the five scenarios (2195 TWh), quantifies the demand-side implications of slower decarbonization.

While this study provides valuable insights, several limitations should be acknowledged. One concern with the model specification is that several explanatory variables constitute major accounting components of final energy consumption. Oil products, natural gas, electricity, biofuels and waste, and coal and peat products jointly represent approximately 94% of historical FEC over the sample period. Therefore, their statistical significance is expected on accounting grounds and should not be interpreted as evidence of independent causal effects. The estimated coefficients mainly reflect conditional associations within the energy-balance structure. These variables instead serve as transition-pathway indicators, allowing the model to translate alternative assumptions regarding fossil-fuel phase-out, electrification, bioenergy deployment, and efficiency improvement into projected FEC. Accordingly, the scenario results are interpreted as conditional projections under specified energy-transition pathways rather than as causal estimates of each carrier's isolated effect on total demand.

Two features of the scenario design bear on this limitation. First, sector coupling is represented only at the energy balance level. The model reallocates demand among energy carriers but does not resolve end-use technologies, such as heat pumps and electric vehicles, through which electrification physically links the heat, transport, and power sectors. Second, the pace of reallocation is set by exogenous milestone assumptions rather than constrained by infrastructural ramp rates such as grid reinforcement or charging network expansion, so the scenarios represent internally consistent transition pathways but do not explicitly evaluate whether the assumed deployment rates are technically or economically feasible. Resolving both would require disaggregating demand by end-use sector and coupling the projections to supply-side and grid integration constraints, which the reduced-form national design adopted here does not accommodate.

While the applied models perform well within the historical data range, their long-horizon projections remain conditional on the assumed future trajectories of the explanatory variables. Distributional shifts associated with deep decarbonization pathways introduce additional uncertainty that constrains forecast reliability. Scheduled structural breaks, such as Germany's coal phase-out, are not modeled as discrete regime changes. Instead, they are reflected indirectly through externally defined predictor trajectories. Capturing these policy transitions explicitly would require models designed to account for regime changes or structural breaks, rather than the static relationships used here. Future research should evaluate the proposed framework using data from additional countries and longer historical datasets as they become available, while also integrating transition dynamics into machine learning models, exploring sector-specific responses to changing energy mixes, and assessing trade-offs between decarbonization ambition, system reliability, and infrastructure investment risks. Such extensions would further support robust, evidence-based energy planning in the context of long-term energy transitions.

5. Conclusion

Long-term energy demand forecasting must account for structural change, policy-driven regime shifts, and short historical records. This study addressed these conditions by benchmarking four models, multiple linear regression (MLR), vector error correction model (VECM), support vector regression (SVR), and artificial neural networks (ANN), to project Germany's final energy consumption (FEC) through 2050. Particle swarm optimization (PSO) and Bayesian optimization (BO) were used to tune the machine learning models, and the comparison prioritized out-of-sample performance, generalization, and long-horizon extrapolation rather than in-sample fit.

Model accuracy and long-horizon stability diverged. The SVR with a radial basis function (RBF) kernel achieved the highest predictive accuracy on the held-out test set (test MAE $= 9.19$, $R^2 = 0.974$) but extrapolated to an implausible increase exceeding the historical maximum, reflecting its sensitivity to nonlinear interactions beyond the observed data range. The ANN attained comparable test accuracy (MAE $= 10.05$, $R^2 = 0.965$) while demonstrating stronger generalization and producing physically plausible long-term projections, and was therefore adopted for the base-case and scenario analyses. MLR provided transparency and interpretability but at lower predictive accuracy, whereas VECM forecasts proved sensitive to structural instability and the assumed cointegration rank, yielding qualitatively different long-term trajectories under small-sample estimation.

The ANN base-case projection indicates that FEC declines by approximately 12% by 2050 relative to 2008, driven by fuel substitution and efficiency improvements. Across the five transition scenarios, projected 2050 FEC ranges from 1476 TWh to 2195 TWh, illustrating the substantial influence of transition assumptions on long-term demand projections. Only the efficiency- and sufficiency-oriented pathway approaches the required scale of demand reduction, yet at approximately 39% below 2008 it still falls short of the national target of a 45% reduction by 2045. These results suggest that, under the assumptions considered in this study, fuel substitution alone is insufficient to achieve the targeted demand reductions and would need to be complemented by stronger efficiency improvements and broader demand-side measures.

The findings should be interpreted in light of several limitations. The short annual record (30 observations) limits the statistical power of cointegration testing and the stability of VECM inference. Highly flexible models can amplify trends when extrapolated beyond the historical data range, as illustrated by the SVR results. In addition, several

predictors constitute major accounting components of FEC, meaning that their statistical importance primarily reflects the underlying energy-balance structure rather than independent causal effects. Finally, all long-term projections remain conditional on the assumed trajectories of the explanatory variables and should therefore be interpreted as scenario-dependent forecasts rather than unconditional predictions.

Future work could integrate transition dynamics directly within machine learning models rather than imposing them through externally projected inputs, disaggregate demand by end-use sector to better represent heterogeneous electrification pathways, and couple demand projections with supply-side and grid-integration constraints. Extending the framework to additional countries and longer historical datasets would further strengthen its generalizability. The annual resolution adopted here constrains the number of observations available for estimation; higher-frequency data, where available, could enlarge the training sample and capture seasonal dynamics, provided appropriate methods are adopted to account for regime heterogeneity and short-run variability. Overall, this study demonstrates that combining systematic hyperparameter optimization with comparative model evaluation provides a robust framework for long-term energy demand forecasting and offers valuable support for evidence-based energy planning under conditions of structural energy transition.

Nomenclature

	Symbols
$\boldsymbol{y}$	Dependent variable (e.g., Final Energy Consumption, FEC)
$\boldsymbol{x}$	Independent variable (predictor)
$\boldsymbol{\beta}$	Regression coefficients (MLR)
$\boldsymbol{\varepsilon}$	Error term
$\boldsymbol{\alpha}$	Adjustment speed in VECM (in the error correction term)
$\boldsymbol{\Gamma}$	Short-term dynamic coefficients matrix (VECM)
$\boldsymbol{\Pi}$	Long-term impact matrix (VECM)
$\boldsymbol{\beta'}$	Cointegration vector transpose (VECM)
$\boldsymbol{det}$	Deterministic term (i.e., seasonality & structural dummies)
$\boldsymbol{\Delta y_t}$	First difference in variable y
$\boldsymbol{p}$	The lag order
$\boldsymbol{w}$	Weight vector, coefficient (SVR, ANN, MLR)
$\boldsymbol{C}$	Regularization parameter (SVR)
ξ	Slack variable
$\boldsymbol{\varepsilon}$	Insensitive loss margin in SVR
$\boldsymbol{\gamma}$	Kernel parameter (SVR, RBF)
$\boldsymbol{\sigma}$	Standard deviation (RBF kernel)
K	Kernel
h	Hidden layer output
$\boldsymbol{\varphi}$	Activation function (ANN)
$\boldsymbol{v}$	Hidden layer weight of ANN
$\boldsymbol{b}$, c	Bias term in ANN
$\boldsymbol{x_{i.t}^d}$	PSO particle position
$\boldsymbol{v_{i.t}^d}$	PSO particle velocity
$\boldsymbol{\omega}$	Inertia weight (PSO)
$\boldsymbol{c_1 \& c_2}$	Cognitive and social component of PSO
$\boldsymbol{p_{i.t}^d}$	PSO particle personal best
$\boldsymbol{p_{l.t}^d}$	PSO global best
$\boldsymbol{r_1 \& r_2}$	Random numbers, stochastic component from (0−1) in PSO
$\boldsymbol{\Delta}$	Change
$\boldsymbol{I}$	Integration order
$\boldsymbol{f(x)}$	Function of x
$\boldsymbol{\mu(x)}$	Mean prediction at x (BO)
$\boldsymbol{exp}$	Exponential function
$\boldsymbol{H_0}$	The null hypothesis
	Subscripts/Superscripts
$\boldsymbol{t}$	Time index (e.g., y_t)
$\boldsymbol{i,j}$	Indices for data points or dimensions
$\boldsymbol{d}$	Dimension index (PSO)
$\boldsymbol{p}$	Lag order (e.g., y_{t-1}) in VECM
$*$	Slack variable falls below the lower ε boundary (SVR, ξ)
	Acronyms and Abbreviations
FEC	Final Energy Consumption
MLR	Multiple Linear Regression
VECM	Vector Error Correction Model
SVM	Support Vector Machine
SVR	Support Vector Regression
ANN	Artificial Neural Network
PSO	Particle Swarm Optimization
BO	Bayesian Optimization
RBF	Radial Basis Function (kernel)
MAE	Mean Absolute Error
MSE	Mean Squared Error
R^2	Coefficient of Determination
GDP	Gross Domestic Product
POP	Population
OP	Oil Products
NG	Natural Gas
EL	Electricity
EE	Energy Efficiency
BIO	Biofuels and Waste
CP	Coal and Peat
RCL	Residential Consumption Level
ADF	Augmented Dickey-Fuller (test)
VAR	Vector Autoregression
ECT	Error Correction Term
CM	Correlation Matrix
LSTM	Long Short-Term Memory
BNetzA	Federal Network Agency (Germany)
BMWK	Federal Ministry for Economic Affairs and Climate Action (Germany)
GHG	Greenhouse Gases
g CO_2eq	Grams of Carbon dioxide equivalent
MJ	Megajoules
min	Minimize
ML	Machine Learning
DL	Deep Learning
STLF	Short-Term Electricity Load Forecasting
ANFIS	Adaptive Neuro-Fuzzy Inference Systems
MRGNN	Multi-Resolution Graph Neural Networks
GA	Genetic Algorithm
SSA	Sparrow Search Algorithm
OLS	Ordinary Least Squares
EI	Expected Improvement
CE	Cointegrating equation
LSSVR	Least-Squares Support Vector Regression
PCA	Principal Component Analysis
MAPE	Mean Absolute Percentage Error

CRediT authorship contribution statement

Movsessian Maria: Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization. **Munzer Osman:** Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Claude (Anthropic) in order to assist with language editing and improving readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of Competing Interest

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

Data Availability

Data will be made available on request.

References

Abdullah, A.S., et al., 2024. Application of machine learning modeling in prediction of solar still performance: A comprehensive survey. Results Eng. 21. https://doi.org/10.1016/j.rineng.2024.101800.

Bakare, M.S., Abdulkarim, A., Shuaibu, A.N., Muhamad, M.M., 2024. A hybrid long-term industrial electrical load forecasting model using optimized ANFIS with gene expression programming (Jun). Energy Rep. 11, 5831–5844. https://doi.org/10.1016/j.egyr.2024.05.045.

Barber, R.F., Candès, E.J., Ramdas, A., Tibshirani, R.J., 2021. Predictive inference with the jackknife+ (Feb). Ann. Stat. 49 (1), 486–507. https://doi.org/10.1214/20-AOS1965.

Bertelsmann Stiftung (Hrsg.), *Wer, wo, wie viele?-Bevölkerungs entwicklung in deutschen Kommunen bis 2040. [Who, where, how many? Population development in German municipalities up to 2040]*. Verlag Bertelsmann Stiftung, 2025. [Online]. Available: www.bertelsmann-stiftung.de/verlag.

Bischl, B., et al., 2023. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges (Mar). ' Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 13 (2). https://doi.org/10.1002/widm.1484.

Bruns, S.B., Stern, D.I., 2019. Lag length selection and p-hacking in Granger causality testing: prevalence and performance of meta-regression models (Mar). Empir. Econ. 56 (3), 797–830. https://doi.org/10.1007/s00181-018-1446-3.

Bruno Burger, Prof. Dr. 'Stromerzeugung in Deutschland im Jahr 2023', Freiburg, 2024. [Online]. Available: https://www.energy-charts.info.

H.G. Buttermann, *AGEB AG Energiebilanzen e. V., 2024. Energy consumption in Germany* 2024. Available: https://ag-energiebilanzen.de/wp-content/uploads/AGEB_Jahresbericht2024_20250815_en.pdf.

Canepa, A., 2022. Small Sample Adjustment for Hypotheses Testing on Cointegrating Vectors (Jan). J. Time Ser. Econ. 14 (1), 51–85. https://doi.org/10.1515/jtse-2020-0044.

Chen, P., 2024. Vector Error Correction Models with Stationary and Nonstationary Variables (Feb). Econ. Anal. Lett. 3 (2), 37–48. https://doi.org/10.58567/eal03020004.

R.B. Christianson and R.B. Gramacy, 'Robust expected improvement for Bayesian optimization', Aug. 2023, [Online]. Available: http://arxiv.org/abs/2302.08612.

H. Dang Nguyen, D. Hägele, J. Lißner, and T. Munz, 'Visual Exploration for Deep Learning Models and Trainings for Microstructure Data', 2022.

Deloitte, *Öko-Institut, Sustainability & Climate: Natural Gas Demand Outlook to 2050, Deloitte and Öko-Institut,* 2023. *Available:* https://www.deloitte.com/content/dam/assets-zone2/de/de/docs/about/2024/Natural_Gas_Demand_Outlook_to_2050_Deloitte_%C3%96ko-Institut.pdf.

Du, K.L., Jiang, B., Lu, J., Hua, J., Swamy, M.N.S., 2024. Exploring Kernel Machines and Support Vector Machines: Principles, Techniques, and Future Directions (Dec). Mathematics 12 (24). https://doi.org/10.3390/math12243935.

Đukanović, M., Kaščelan, L., Vuković, S., Martinović, I., Ćalasan, M., 2023. A machine learning approach for time series forecasting with application to debt risk of the Montenegrin electricity industry (Sep). Energy Rep. 9, 362–369. https://doi.org/10.1016/j.egyr.2023.05.240.

Eurostat, 'Population projections in the EU Statistics Explained', 2023. [Online]. Available: https://ec.europa.eu/eurostat/statisticsexplained/.

Federal Ministry for Economic Affairs and Climate Action (BMWK), *System Development Strategy 2024, BMWK, Berlin*, 2024. Official BMWK PDF. Available: https://www.bundeswirtschaftsministerium.de/Redaktion/EN/Publikationen/Klimaschutz/2024-system-development-strategy.pdf?_blob=publicationFile&v=2.

Federal Ministry for the Environment, 'Nature Conservation, Building and Nuclear Safety (BMUB). Climate Action Plan 2050: Principles and goals of the German government's climate policy', 2016. [Online]. Available: www.bmub.bund.de/english.

Federal Network Agency (Bundesnetzagentur), 'Approval of the Scenario Framework 2025–2037/2045 for the Electricity Network Development Plan', Bonn, 2025. [Online]. Available: https://www.netzentwicklungsplan.de/sites/default/files/2025-05/250430_Genehmigung_Szenariorahmen_2025.pdf.

Federal Office for Agriculture and Food (BLE), 'Evaluations- und Erfahrungsbericht für das Jahr 2023: Biomassestrom-Nachhaltigkeitsverordnung, Biokraftstoff-Nachhaltigkeitsverordnung', Bonn, 2024. [Online]. Available: https://www.ble.de/SharedDocs/Downloads/DE/Klima-Energie/Nachhaltige-Biomasseherstellung/Evaluationsbericht_2023.pdf?_blob=publicationFile&v=2.

Federal Republic of Germany, 'Bundes-Klimaschutzgesetz (KSG) *[Federal Climate Change Act], Bundesgesetzblatt I (BGBl. I), p. 2513',* 2019*, last amended 15 Jul. 2024 (BGBl. 2024 I No. 235)*. https://www.gesetze-im-internet.de/ksg/KSG.pdf.

Federal Republic of Germany, 'Gesetz zur Reduzierung und zur Beendigung der Kohleverstromung (Kohleverstromungsbeendigungsgesetz, KVBG) [Coal Exit Act], Bundesgesetzblatt I', 2020. [Online]. Available: https://www.gesetze-im-internet.de/kvbg/.

Federal Republic of Germany, 'Law to Increase Energy Efficiency in Germany (Energy Efficiency Act – EnEfG), Bundesgesetzblatt Teil I (BGBl. I), No. 309', Bonn, 2023.

Fodstad, M., et al., 2022. Next frontiers in energy system modelling: A review on challenges and the state of the art (May). Renew. Sustain. Energy Rev. 160. https://doi.org/10.1016/j.rser.2022.112246.

Fumagalli, F., Muschalik, M., Hüllermeier, E., Hammer, B., 2023. Incremental permutation feature importance (iPFI): towards online explanations on data streams (Dec). Mach. Learn. 112 (12), 4863–4903. https://doi.org/10.1007/s10994-023-06385-y.

Gao, J., Peng, B., Yan, Y., 2025. Time-varying vector error-correction models: Estimation and inference (Sep). J. Econom. 251. https://doi.org/10.1016/j.jeconom.2025.106035.

Granger, C.W.J., 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods (Aug). Econometrica 37 (3), 424. https://doi.org/10.2307/1912791.

Guangxiao, L., Yilu, Y., Zongjie, L., Zhaohui, H., Changyun, L., 2025. Electricity demand forecasting and power supply planning under carbon neutral targets (Jun). Energy Rep. 13, 2740–2751. https://doi.org/10.1016/j.egyr.2025.02.015.

Hansen, B.E., 2022. Econometrics. Princeton University Press, *Princeton.* https://api.pageplace.de/preview/DT0400.9780691236155_A42917420/preview-9780691236155_A42917420.pdf (Available).

International Monetary Fund (IMF), 'World Economic Outlook Update Global Growth: Divergent and Uncertain', 2025. [Online]. Available: https://www.imf.org/-/media/files/publications/weo/2025/october/english/text.pdf.

Jain, A., Gupta, S.C., 2024. Evaluation of electrical load demand forecasting using various machine learning algorithms. Front. Energy Res. 12. https://doi.org/10.3389/fenrg.2024.1408119.

Johansen, S., 1991. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models (Nov). Econometrica 59 (6), 1551. https://doi.org/10.2307/2938278.

Joseph, V.R., Vakayil, A., 2022. SPlit: An Optimal Method for Data Splitting. Technometrics 64 (2), 166–176. https://doi.org/10.1080/00401706.2021.1921037.

Kluczek, A., Woźniak, A., Żegleń, P., 2025. National diversity in European energy policy: Analyzing dependencies of changes in energy prices, climate regulations, and technological innovations on economic implications (Nov). Energy Strategy Rev. 62. https://doi.org/10.1016/j.esr.2025.101886.

Łasak, K., Lont, J., 2020. Observation Driven Long Run Equilibria (Feb). Comput. Econ. 55 (2), 551–575. https://doi.org/10.1007/s10614-019-09903-0.

Li, J., Lei, Y., Yang, S., 2022. Mid-long term load forecasting model based on support vector machine optimized by improved sparrow search algorithm (Aug). Energy Rep. 8, 491–497. https://doi.org/10.1016/j.egyr.2022.02.188.

M. Lizarazo López, T. Petersen, S. Kaniovski, and T. Url, 'Macroeconomic effects of demographic aging', 2022. [Online]. Available: www.demografischer-wandel.de.

Lyhagen, J., Baris-Tüzemen, Ö., 2025. Estimating cointegrating models in the presence of additive outliers. Commun. Stat. Simul. Comput. https://doi.org/10.1080/03610918.2025.2577404.

Mo, S., et al., 2025. MRGNN: Multi-resolution graph neural network for electricity load forecasting (Jun). Energy Rep. 13, 4624–4634. https://doi.org/10.1016/j.egyr.2025.03.056.

Montesinos López, O.A., López, A.M., Crossa, J., 2022. Multivariate Statistical Machine Learning Methods for Genomic Prediction (Jan). Springer, pp. 1–691. https://doi.org/10.1007/978-3-030-89010-0.

Mystakidis, A., Koukaras, P., Tsalikidis, N., Ioannidis, D., Tjortjis, C., 2024. Energy Forecasting: A Comprehensive Review of Techniques and Technologies (Apr). Energ. (Basel) 17 (7). https://doi.org/10.3390/en17071662.

Ozdemir, G., 2024. Long-term electrical energy demand forecasting by using artificial intelligence/machine learning techniques (Aug). Electr. Eng. 106 (4), 5229–5251. https://doi.org/10.1007/s00202-024-02364-1.

Ozerdem, O.C., Olaniyi, E.O., Oyedotun, O.K., 2017. Short term load forecasting using particle swarm optimization neural network. Procedia Computer Science. Elsevier B. V, pp. 382–393. https://doi.org/10.1016/j.procs.2017.11.254.

Pinheiro, J.M.H., et al., 2025. The Impact of Feature Scaling in Machine Learning: Effects on Regression and Classification Tasks. IEEE Access. 13, 199903–199931. https://doi.org/10.1109/ACCESS.2025.3635541.

Sensfuß, F., et al., 2021. Long-term Scenarios for the Transformation of the Energy System in Germany III - Summary Report: 3 Main Scenarios. Fraunhofer ISI Consent. ifeu TU Berl. Karlsr. Ger. https://www.langfristszenarien.de/enertile-explorer-wAssets/docs/LFS3-Kurzbericht_EN_formatiert-20211011.pdf (Available).

Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., De Freitas, N., 2016. Taking the human out of the loop: A review of Bayesian optimization (Jan). Proc. IEEE 104 (1), 148–175. https://doi.org/10.1109/JPROC.2015.2494218.

Shojaie, A., Fox, E.B., 2022. Granger Causality: A Review and Recent Advances. Annu. Rev. Stat. Appl. 37, 20. https://doi.org/10.1146/annurev-statistics-040120.

Stermieri, L., Kober, T., McKenna, R., Schmidt, T.J., Panos, E., 2023. Impacts of digitalization and societal changes on energy transition: a novel socio-techno-economic energy system model (Nov). Energy Strategy Rev. 50. https://doi.org/10.1016/j.esr.2023.101224.

Stock, J.H., Watson, M.W., 2020. Introduction to Econometrics, fourth edition., global edition. Pearson Education Limited, *Harlow*.

Tharwat, A., 2019. Parameter investigation of support vector machine classifier with kernel functions (Dec). Knowl. Inf. Syst. 61 (3), 1269–1302. https://doi.org/10.1007/s10115-019-01335-4.

Twumasi, E., Frimpong, E.A., Prah, N.K., Gyasi, D.B., 2024. A novel improvement of particle swarm optimization using an improved velocity update function based on local best murmuration particle (Oct). J. Electr. Syst. Inf. Technol. 11 (1). https://doi.org/10.1186/s43067-024-00168-8.

UNdata, 'United Nations (UN) statistical system', n.d. [Online]. Available: https://data.un.org/Host.aspx?Content=About.

Wen, J., Li, Y., Wang, J., Sun, W., 2022. Nonstationary Process Monitoring Based on Cointegration Theory and Multiple Order Moments (Jan). Processes 10 (1). https://doi.org/10.3390/pr10010169.

Wu, J., Wang, Y.G., 2023. A working likelihood approach to support vector regression with a data-driven insensitivity parameter (Mar). Int. J. Mach. Learn. Cybern. 14 (3), 929–945. https://doi.org/10.1007/s13042-022-01672-x.

Yang, Y., Han, L., Wang, Y., Wang, J., 2022. China's Energy Demand Forecasting Based on the Hybrid PSO-LSSVR Model (Jan). Wirel. Commun. Mob. Comput. 2022, 1–12. https://doi.org/10.1155/2022/7584646.

Yingnan, W., Xiaowei, W., 2025. A hybrid power load forecasting model based on evolutionary strategy and long short term memory (Dec). Energy Rep. 14, 845–853. https://doi.org/10.1016/j.egyr.2025.05.041.

Yu, M., Zhou, Y.N., Wang, Q., Yan, F., 2024. Extrapolation validation (EV): a universal validation method for mitigating machine learning extrapolation risk (Apr). Digit. Discov. 3 (5), 1058–1067. https://doi.org/10.1039/d3dd00256j.

Zivich, P.N., Naimi, A.I., 2025. A primer on neural networks (Jun). Am. J. Epidemiol. 194 (6), 1473–1475. https://doi.org/10.1093/aje/kwae380.