Ioannis Skevas* and Theologos Dergiades

Efficiency Analysis and Bayesian Neural Networks

https://doi.org/10.1515/snde-2025-0091
Received July 2, 2025; accepted December 18, 2025; published online January 8, 2026

Abstract: A key limitation of traditional efficiency measurement is the inability of typical functional forms to capture non-linearities. This study proposes an alternative framework for efficiency analysis by integrating a Bayesian neural network with a generalized true random-effects Stochastic Frontier Analysis (SFA) model. In doing so, complex functional forms are better approximated, while both time-varying and time-invariant inefficiencies are considered for. The neural network SFA model is empirically compared with the conventional SFA model using a Cobb-Douglas specification. The results highlight that the neural network model better captures underlying non-linearities and provides a more accurate assessment of inefficiency. Bayes factors provide strong evidence in favor of the neural network model. These initial findings signal the potential of neural networks to enhance the precision and flexibility of efficiency analysis.

Keywords: Bayesian neural network; stochastic frontier analysis; generalized true random effects

JEL Classification: C11; C33; C45; D22

1 Introduction

Machine learning techniques are essential tools for practitioners analyzing real-world data, particularly for revealing complex patterns that might otherwise be missed (Hauzenberger et al. 2025). These techniques are highly relevant across various sectors characterized by complex relationships, including, among others, agriculture, healthcare, manufacturing, and energy. This relevance is especially critical in the context of inefficiency measurement, as the precision of performance measures that guide firms, policymakers, and other stakeholders can be compromised by the limitations of conventional modeling approaches (Giannakas et al. 2003). Specifically, when applying the stochastic frontier analysis (SFA) technique introduced by Aigner et al. (1977) and Meeusen and van den Broeck (1977) to quantify inefficiency, researchers often rely on standard functional forms (i.e. Cobb-Douglas) to model the relationship between inputs and outputs. While standard functional forms are widely used for their simplicity and convenience, they may not adequately capture the highly non-linear relationships between inputs and outputs.

In this study, we examine the consequences of ignoring non-linearities in the production frontier when estimating time-invariant and time-varying inefficiencies. To do so, we (1) specify an SFA model that distinguishes between time-invariant and time-varying inefficiency components, (2) model the deterministic part of the frontier using a neural network to capture non-linearities in the input−output relationship, and (3) compare the results with those obtained from a conventional SFA model that assumes a linear Cobb−Douglas functional form for the deterministic part of the frontier.

***Corresponding author: Ioannis Skevas**, Assistant Professor, Department of International & European Studies, University of Macedonia, 156 Egnatia Street, PC 54636, Thessaloniki, Greece, E-mail: gskevas@uom.edu.gr. https://orcid.org/0000-0003-4595-4396
Theologos Dergiades, Assistant Professor, Department of International & European Studies, University of Macedonia, 156 Egnatia Street, PC 54636, Thessaloniki, Greece, E-mail: dergiades@uom.edu.gr. https://orcid.org/0000-0003-3294-7787

Accounting for frontier non-linearities is particularly important in sectors where input–output relationships are complex and influenced by external conditions. In the energy sector, which serves as the empirical context of this study, the marginal productivity of inputs like labor, capital, and water can vary significantly depending on weather conditions, resource availability, and the regulatory framework. Such non-linear interactions may lead to sizable deviations from the predictions made by standard functional forms, highlighting the need for more flexible modeling approaches. Neural networks are powerful machine learning tools that have proven to be highly effective as function approximators, learning to map inputs to outputs by adjusting the network's weights based on the available data (Pendharkar 2023). In particular, the network adjusts its parameters, enabling it to "learn" the underlying function that relates inputs to outputs, even when this function is complex or highly non-linear (Azadeh et al. 2010). This attribute has made neural networks attractive for studies such as those by Azadeh et al. (2010), Konstantakis et al. (2025), Liao et al. (2007), Pendharkar (2023), Wang (2003), and Xuan (2018), which have applied them in SFA to approximate the unknown functional form of the frontier. Additionally, neural networks have been employed to address uncertainty in the relationship between the inefficiency term and socio-economic characteristics, as highlighted in Tsionas et al. (2023).

To date, most of the studies in the field have not adequately addressed the panel structure of their data, particularly with regard to differentiating between time-varying and time-invariant inefficiency components, as suggested by Kumbhakar et al. (2014) and Tsionas and Kumbhakar (2014) in their generalized true random effects model. Additionally, the potential consequences of neglecting the presence of non-linearities between inputs and outputs within both time-invariant and time-varying inefficiency scores remain unexplored. In this study, we address these gaps by integrating a generalized true random effects SFA model with a Bayesian neural network to simultaneously capture both time-varying and time-invariant inefficiencies while accounting for potential non-linear relationships. We also estimate inefficiencies using the conventional Cobb-Douglas specification for the frontier function and compare these estimates with those derived from the proposed neural network model. To illustrate the empirical implications of our approach, we apply both models to a panel dataset from the energy sector. By comparing the results of the neural network-based model with those from the conventional Cobb–Douglas specification, we highlight the practical significance of adopting a more flexible functional form in frontier analysis.

The remainder of the paper is structured as follows. In the next section, we introduce the proposed generalized true random effects SFA model along with the specified neural network structure and estimation procedure. We also present the conventional SFA model based on a Cobb–Douglas functional form and outline the approach used to compare the two models. This is followed by a description of the dataset used in the analysis. The subsequent section presents the results obtained from the proposed neural network-based SFA model and compares them to those derived from the conventional Cobb–Douglas specification. The final section offers concluding remarks.

2 Models and Estimation

2.1 Neural Network SFA Model

The generalized true random effects SFA model for firm i at time t is written as:

$$y_{t,i} = \mu_{t,i} + a_i + \upsilon_{t,i} - \eta_i - u_{t,i} \tag{1}$$

where $y_{t,i}$ is the log-transformed observed output and $\mu_{t,i}$ denotes the deterministic part of the production frontier. Regarding the random variables, a_i is the firm-specific random effect, capturing unobserved heterogeneity, and is assumed to follow a Normal distribution, $a_i \sim \mathcal{N}(0, \omega^{-1})$. The term $\upsilon_{t,i}$ represents the noise component, accounting for random shocks, and is also assumed to be Normally distributed as $\upsilon_{t,i} \sim \mathcal{N}(0, \tau^{-1})$. The inefficiency components, η_i and $u_{t,i}$ represent time-invariant (thereafter persistent) and time-varying (thereafter transient) inefficiencies, respectively, and are modeled with a Half-Normal distribution, where $\eta_i \sim \mathcal{N}^{+}(0, \psi^{-1})$ and

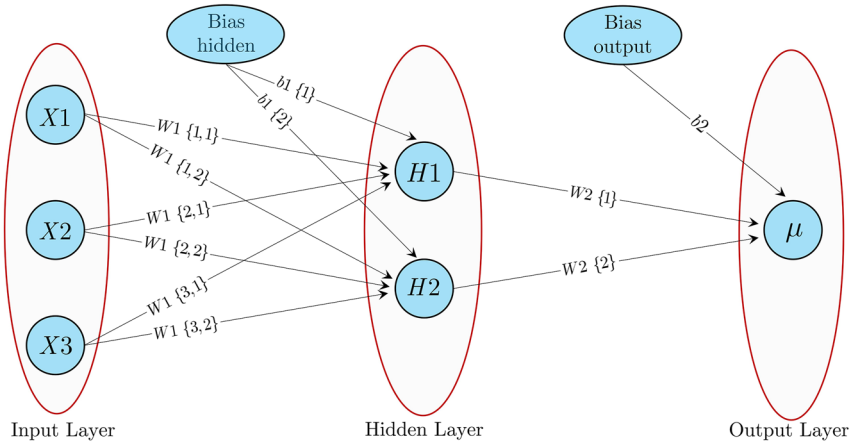

Figure 1: Structure of the neural network.

$u_{t,i} \sim \mathcal{N}^{+}(0, \phi^{-1})$.[1] The deterministic part $\mu_{t,i}$ is modeled using a neural network that captures the potentially complex, non-linear relationship between inputs and output:

$$\mu_{t,i} = \sum_{j=1}^{H} W2_j \cdot f\left(\sum_{k=1}^{K} W1_{k,j} \cdot X_{t,i,k} + b1_j\right) + b2 \tag{2}$$

where $X_{t,i,k}$ represents the k-th log-transformed input variable for firm i at time t, $W1_{k,j}$ are the weights from the input layer to the hidden layer of the neural network, $W2_j$ are the weights from the hidden layer to the output layer, $b1_j$ and $b2$ are the biases for the hidden and output layers, respectively, $f(\cdot)$ is the activation function applied to the weighted sums, and H is the number of hidden neurons.

The structure of a simple neural network designed for a scenario with three input features ($X1$, $X2$, and $X3$), a single hidden layer containing 2 neurons ($H1$ and $H2$), and one output (μ) is illustrated in Figure 1. The network begins with the 3 inputs, which represent the features of the dataset and serve as the starting point for neural computation. These inputs are fed into the hidden layer, where each is multiplied by corresponding weights associated with each connection to the hidden neurons. For each hidden neuron j, the neuron computes a weighted sum of its inputs plus a bias term, expressed as $\sum_{k=1}^{K} W1_{k,j} \cdot X_{t,i,k} + b1_j$, which is then passed through an activation function $f(\cdot)$ to introduce non-linearity into the model. The outputs from the hidden neurons are then linearly combined to produce the final output μ, using another set of weights $W2_j$ and a bias term $b2$. This output represents the neural network's prediction. Bias terms in both the hidden and output layers play a crucial role in shifting the activation functions, allowing the network to better fit the data by accounting for systematic deviations.

Regarding the structure of the neural network in this study, the following decisions are made: 1) the hyperbolic tangent (tanh) function is chosen as the activation function $f(\cdot)$, due to its ability to introduce non-linearity into the model (Azadeh et al. 2007). The tanh function maps inputs to a range between -1 and 1, providing a zero-centered output that facilitates balanced training and improves convergence, 2) a single hidden layer is specified, based on the universal approximation theorem, which states that a neural network with one hidden layer can approximate any continuous function (Wang 2003), 3) the number of neurons in the hidden layer H is determined using the heuristic that suggests H should be approximately 2/3 of the size of the input layer plus the size of the output layer (Heaton 2008). This rule of thumb balances model complexity by reducing the risk of overfitting while maintaining sufficient model capacity. Given that the subsequent application includes 5 independent variables and 1 output variable, H is set equal to four.

1 The terms ω, τ, ψ, and ϕ are the precision (i.e. inverse variance) parameters to be estimated.

2.2 Bayesian Estimation

A Bayesian approach is used to estimate the model, offering key advantages, especially in handling latent variables. This is particularly useful in neural networks, where weights and biases are treated as random variables with prior distributions. Instead of focusing on a single optimal value, Bayesian methods estimate their full posterior distributions, capturing uncertainty and providing a more comprehensive view of plausible parameter values. Bayesian estimation requires specifying the data likelihood and prior distributions for the parameters. The data likelihood is a product of two Normal distributions for a_i and $\upsilon_{t,i}$, and two Half-Normal distributions for η_i and $u_{t,i}$:

$$\begin{aligned} & p(\{y_{t,i}\},\{a_i\},\{\eta_i\},\{u_{t,i}\} \mid \{\mathbf{X}_{t,i}\},\mathbf{W1},\mathbf{W2},\mathbf{b1},b2,\omega,\psi,\phi,\tau) \\ &= \frac{\omega^{N/2}}{(2\pi)^{N/2}} \exp\left\{-\frac{\omega}{2}\sum_{i=1}^{N} a_i^2\right\} \\ &\times \frac{2^N\psi^{N/2}}{(2\pi)^{N/2}} \exp\left\{-\frac{\psi}{2}\sum_{i=1}^{N} \eta_i^2\right\} \\ &\times \frac{2^{NT}\phi^{NT/2}}{(2\pi)^{NT/2}} \exp\left\{-\frac{\phi}{2}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{t,i}^2\right\} \\ &\times \frac{\tau^{NT/2}}{(2\pi)^{NT/2}} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{N}\sum_{t=1}^{T} \upsilon_{t,i}^2\right\} \end{aligned} \tag{3}$$

The prior distributions for the weights, $p(\mathbf{W1})$ and $p(\mathbf{W2})$, and bias terms, $p(\mathbf{b1})$ and $p(b2)$, are Normal with a mean of 0 and a precision of 0.01. Since the variance of a Normal distribution is the inverse of its precision, this corresponds to a variance of 100, allowing for a wide range of possible values and reflecting limited prior information about the network parameters. The precision parameters $p(\omega)$, $p(\psi)$, $p(\phi)$, and $p(\tau)$ follow Gamma distributions. For $p(\omega)$ and $p(\tau)$, both the shape and rate hyperparameters are set to 0.001, resulting in a very flat and uninformative prior. The variance of a Gamma distribution equals the shape parameter divided by the square of the rate parameter, which in this case yields a variance of 1,000, indicating considerable prior uncertainty. In contrast, the priors for $p(\psi)$ and $p(\phi)$, which govern the inefficiency terms, use a shape of 7 and a rate of 0.5. Because the mean of a Gamma-distributed random variable equals the ratio of its shape to its rate, these values imply a prior mean of 14 for the corresponding precision parameters. Given that the mean of a Half-Normal distribution equals $\sqrt{(1/\text{precision})}\sqrt{2/\pi}$, this specification translates into a prior mean inefficiency of approximately 0.2. This value is consistent with the recommendation of Van den Broeck et al. (1994), and ensures the prior is centered around plausible values observed in empirical applications. The posterior distribution, obtained by combining the likelihood and priors, is:

$$\begin{aligned} & p(\{a_i\},\{\eta_i\},\{u_{t,i}\},\mathbf{W1},\mathbf{W2},\mathbf{b1},b2,\omega,\psi,\phi,\tau \mid \{y_{t,i}\},\{\mathbf{X}_{t,i}\}) \\ &\propto p(\{y_{t,i}\},\{a_i\},\{\eta_i\},\{u_{t,i}\} \mid \{\mathbf{X}_{t,i}\},\mathbf{W1},\mathbf{W2},\mathbf{b1},b2,\omega,\psi,\phi,\tau) \\ &\times p(\mathbf{W1}) \times p(\mathbf{W2}) \times p(\mathbf{b1}) \times p(b2) \times p(\omega) \times p(\psi) \times p(\phi) \times p(\tau) \end{aligned} \tag{4}$$

Samples from the posterior distribution are obtained using Markov Chain Monte Carlo simulation.

2.3 Conventional SFA Model and Comparison

To fulfill the main objective of this study, which is to highlight the consequences of ignoring non-linearities in the production frontier when estimating persistent and transient inefficiencies, we also specify a conventional Cobb–Douglas SFA model. This specification serves as a benchmark and is obtained by defining the deterministic component of the frontier $\mu_{t,i}$ as a linear combination of the log-transformed input variables and their associated parameters:

$$
\begin{aligned}
y_{t,i} &= \mu_{t,i} + a_i + v_{t,i} - \eta_i - u_{t,i} \\
\mu_{t,i} &= \sum_{k=1}^{K} X_{t,i,k} \cdot \beta_k
\end{aligned}
\tag{5}
$$

To ensure comparability, all distributional assumptions for the model's random components are kept identical to those used in the neural network SFA model. The same applies to the prior distributions used for Bayesian estimation.

After estimating both the neural network and conventional SFA models, we conduct Bayesian hypothesis testing to compare the resulting inefficiency estimates. This testing procedure is based on the posterior draws of the inefficiency scores. Specifically, we focus on two summary statistics: the means and the standard deviations of the inefficiencies. The procedure involves the following steps: (1) extracting the mean and standard deviation of the inefficiency scores from each posterior draw, (2) sorting these values in ascending order, (3) computing the differences between the two models for each draw, and (4) discarding the lowest and highest 2.5 % of the differences to construct a 95 % credible interval. If this interval includes zero, we do not reject the null hypothesis that the two models yield equivalent inefficiency moments. The null and alternative hypotheses are formally stated as follows:

$$
\begin{aligned}
H_0&: \theta_1 - \theta_2 = 0 \\
H_1&: \theta_1 - \theta_2 \neq 0
\end{aligned}
\tag{6}
$$

where θ_1 and θ_2 denote the posterior means (or standard deviations) of the inefficiency scores from the Cobb–Douglas and neural network SFA models, respectively.

Finally, we compare the competing models based on their marginal log-likelihoods, which are approximated using the Laplace–Metropolis estimator (Lewis and Raftery 1997). The relative support for each model is assessed using Bayes factors, defined as the ratio of the marginal likelihoods of two models:

$$
\mathrm{BF}_{12} = \frac{p(\mathcal{D} \mid M_1)}{p(\mathcal{D} \mid M_2)},
\tag{7}
$$

where $p(\mathcal{D} \mid M_i)$ denotes the marginal likelihood of model M_i given the data $\mathcal{D}$. Assuming equal prior probabilities $p(M_1) = p(M_2) = 0.5$, the posterior probability of model M_i is given by Bayes' theorem:

$$
p(M_i \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid M_i) \cdot p(M_i)}{\sum_{j=1}^{2} p(\mathcal{D} \mid M_j) \cdot p(M_j)}.
\tag{8}
$$

Posterior probabilities are normalized so that they sum to one across models. The model with the highest posterior probability is considered the most strongly supported by the data. In addition, the Deviance Information Criterion (DIC), introduced by Spiegelhalter et al. (2002), is also reported for each model, with lower values indicating a better fit.

3 Data

To empirically compare the neural network SFA model with the conventional SFA model, we use the dataset from Kerstens et al. (2019).[2] The dataset, recorded monthly, includes the output and inputs of 16 Chilean hydroelectric power plants over the entire year of 1997, comprising a total of 192 observations. The output variable is the electricity generated (Q) in Gigawatt-hours (GWh), while the three input variables are: (a) the water volume (M) in thousands of cubic meters, (b) the labor quantity (L), measured by the total number of full-time employees, and (c) the value of capital (K) in real Chilean pesos. The raw observations for the output variable (Q) and the three input variables (M, L, and K) are illustrated analytically in Figures 2, 3, 4, and 5, respectively. In addition

2 Kerstens et al. (2019) use a subset of a larger dataset originally constructed by Atkinson and Halabí (2005) and later made publicly accessible by Atkinson and Dorfman (2009).

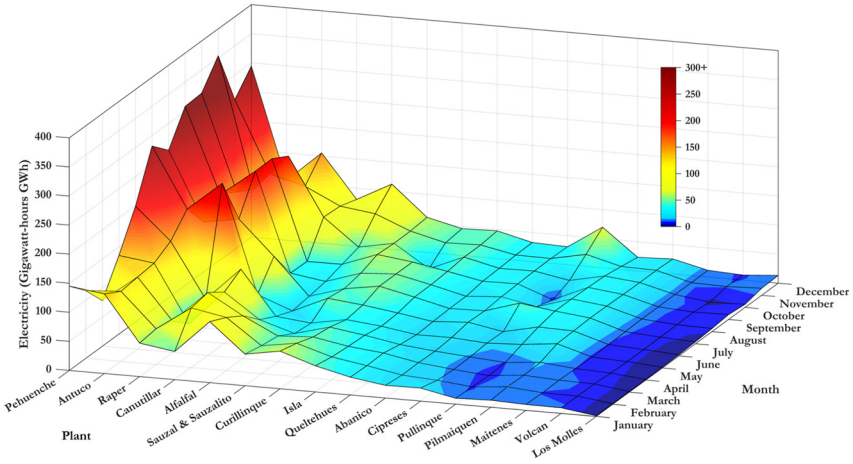

Figure 2: Electricity generated.

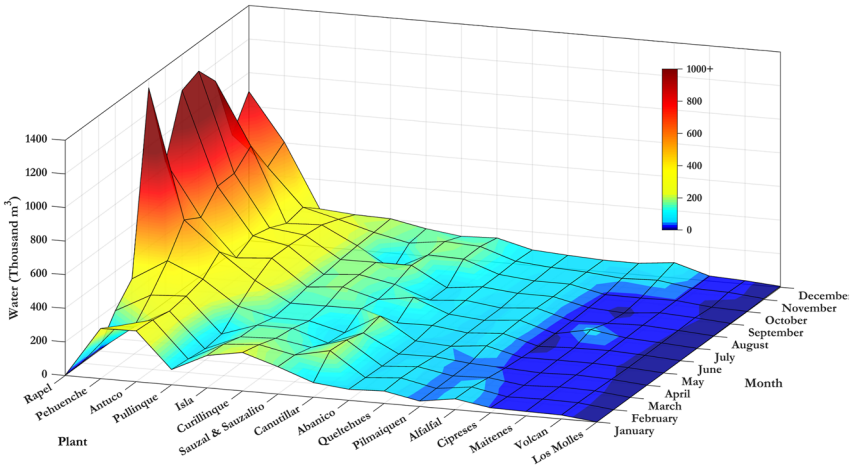

Figure 3: Water volume.

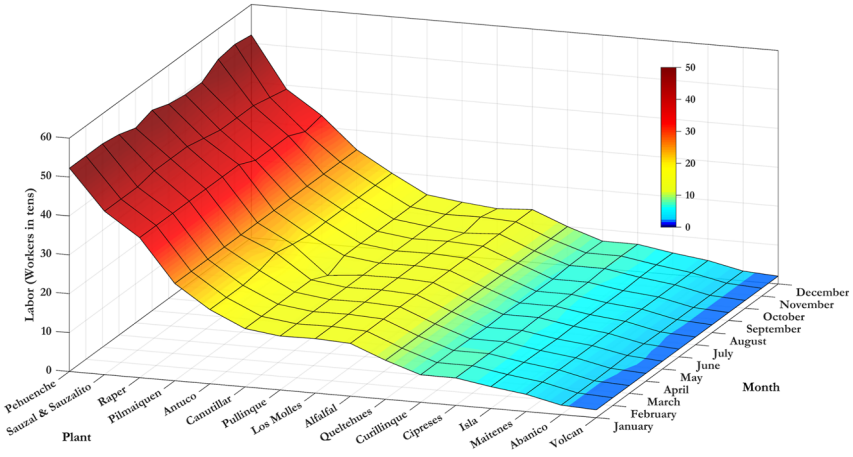

Figure 4: Labor quantity.

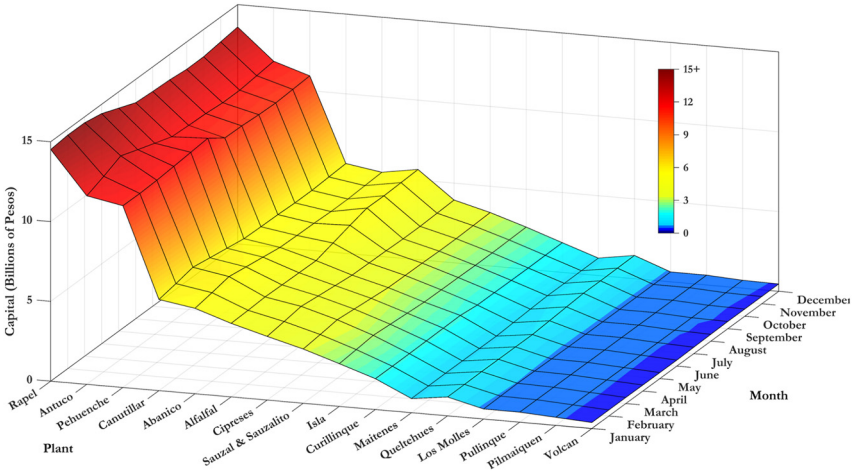

Figure 5: Value of capital.

to these inputs, the model specification includes an intercept and a monthly time trend to capture temporal variation throughout the year.

4 Results

The utilized sampling scheme involves a total of 700,000 iterations, with the first 300,000 discarded as a burn-in phase to allow the Markov Chains to stabilize. After burn-in, 400,000 iterations are used, but only every 10^{th} iteration is retained, resulting in 40,000 posterior samples. Thinning helps reduce autocorrelation between consecutive samples, ensuring more reliable parameter estimates. The parameter estimates from both the neural network SFA model and the conventional SFA model are presented in the Appendix in Tables A.1 and A.2, respectively. In addition to standard statistics (i.e. mean, standard deviation, and 95 % credible interval), the corresponding Geweke's diagnostic and Monte Carlo standard error (MCSE) are also reported to evaluate the convergence and accuracy of the MCMC simulations. The Geweke's diagnostic compares the means of the first and last portions of the Markov chain; when the chain has converged to its stationary distribution, these means should not differ systematically. Hence, values close to zero indicate satisfactory convergence. The MCSE quantifies the uncertainty associated with estimating the posterior mean from a finite number of draws and provides an indication of the effective sample size. Small MCSE values suggest that the posterior estimates are numerically stable and that sampling autocorrelation is minimal. Together, these diagnostics confirm that the MCMC chains have mixed well and that the reported parameter estimates are based on reliable and well-converged posterior samples.

The results from the neural network SFA model, presented in Table A.1, indicate that several weights $W1$, which determine how each input variable contributes to the activation of the hidden neurons, exhibit strong posterior signals with narrow credible intervals that exclude zero. In particular, weights associated with the labor and capital inputs contribute significantly to the non-linear structure of the estimated production frontier. For example, the parameter $W1_{3,1}$, linking labor to the first hidden neuron, has a posterior mean of 7.781 with a 95 % credible interval well above zero, indicating a strong positive contribution. The weights in the output layer ($W2$) determine how each hidden neuron contributes to the final prediction of the production frontier. Among them, $W2_2$ stands out with a posterior mean of 6.816 and a 95 % credible interval of $[3.771, 11.689]$, suggesting that the second hidden neuron plays a substantial and statistically significant role in shaping the model's output.

The bias terms $b1_1$ and $b1_2$ are large in magnitude and significant, highlighting their role in shifting the activation functions and improving model flexibility. Overall, the estimated network parameters confirm that the neural network SFA model effectively captures complex patterns between the input variables and the output,

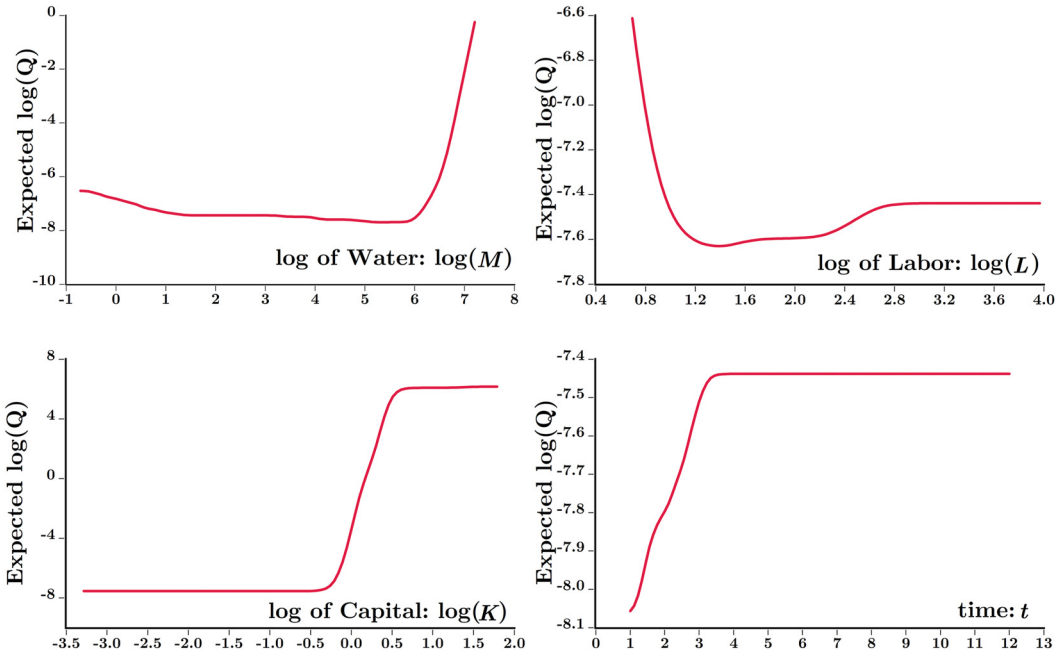

Figure 6: Expected output versus inputs and time (M, L, K and t).

beyond what a linear specification can achieve. In contrast, the Cobb–Douglas SFA model presented in Table A.2 imposes constant elasticities, with each input affecting output in a fixed, proportional manner. Although all estimated coefficients in the Cobb–Douglas model are significant, the linear functional form limits its ability to represent more complex or varying relationships among inputs and output.

The relationships between each input variable and time with the expected output resulted from the neural network SFA model are illustrated in Figure 6. These relationships are shaped by the *tanh* activation function in the neural network's hidden layer, resulting in S-shaped curves. Initially, increasing water quantity has little effect on the expected output, but beyond a certain threshold, a sharp increase occurs, suggesting that water quantity becomes a major factor driving production. For labor, lower values correspond to reduced output, and after a certain point, further increases in labor have minimal positive impact. Similarly, capital has little effect on the expected output when it is below a specific level, but once it crosses this threshold, it significantly boosts production, though additional increases in capital yield diminishing effects. Finally, the expected output increases rapidly in the early periods of time but stabilizes as time progresses. These non-linear relationships, including threshold effects, cannot be captured by standard linear functional forms like the Cobb-Douglas, as suggested by the constant positive elasticities shown in Table A.2 of the Appendix. This highlights the importance of using neural networks to uncover more complex input-output relationships.

Moreover, Figure 7 presents the density plots of persistent and transient inefficiencies for both the neural network SFA model and the conventional SFA model with a Cobb-Douglas functional form. In the left plot, which shows persistent inefficiency, the average for the neural network SFA model is slightly lower at 0.146 compared to 0.159 in the conventional SFA model. However, more substantial differences are observed in the right plot, which shows the transient inefficiency densities for the two models. Both distributions are right-skewed, but transient inefficiencies are significantly higher in the conventional SFA model. The average transient inefficiency is estimated at 0.235 for the neural network SFA model, while it is notably larger at 0.467 for the conventional SFA model. Additionally, the conventional model exhibits more outliers in the distribution of transient inefficiencies, suggesting a greater spread in inefficiency across units. The above differences in the inefficiency scores across the two models are empirically tested for significance using the procedure outlined in Subsection 2.3. Figure 8 presents histograms of the differences in the mean persistent and transient inefficiencies, highlighting the posterior mean differences along with their corresponding 95 % credible intervals. Focusing on Figure 8a (left

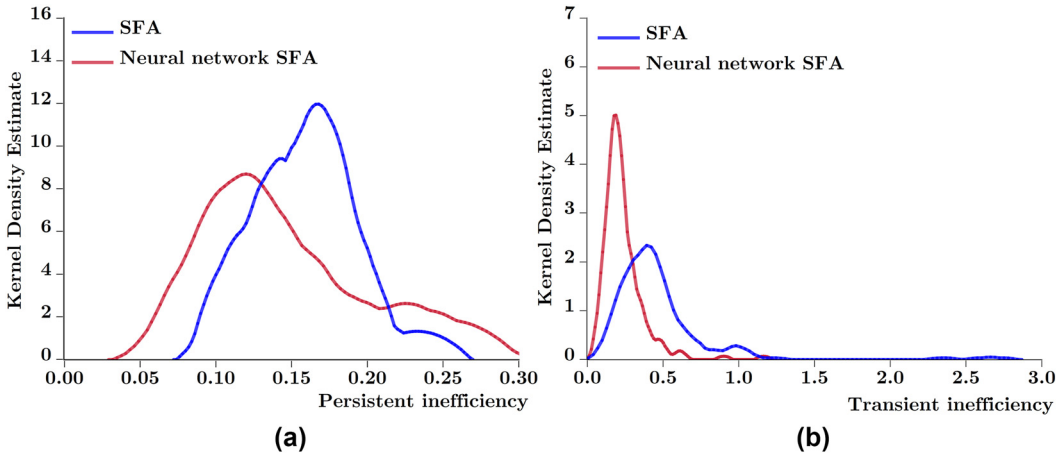

Figure 7: Kernel densities of persistent (η_i) and transient ($u_{t,i}$) inefficiencies.

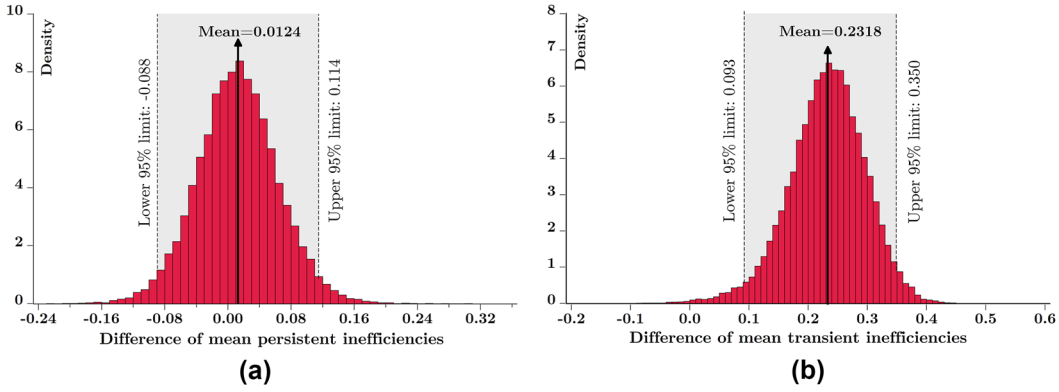

Figure 8: Histograms of differences in mean inefficiencies.

panel), the difference in the mean persistent inefficiencies between the two models is relatively small, at 0.012. This is consistent with the previously reported averages: 0.159 for the conventional Cobb–Douglas model and 0.146 for the neural network SFA model. The 95 % credible interval for the difference, [−0.088, 0.114], includes zero, indicating that the difference is not significant. In contrast, Figure 8b (right panel) shows a substantial difference in the mean transient inefficiencies. The average transient inefficiency is 0.467 under the Cobb–Douglas model and 0.235 under the neural network SFA model, yielding a posterior mean difference of approximately 0.232. The corresponding 95 % credible interval, [0.093, 0.350], lies entirely above zero, providing strong evidence that the models yield significantly different estimates of transient inefficiency. This result suggests that the conventional Cobb–Douglas specification inflates the level of transient inefficiency, likely due to its inability to flexibly capture the underlying input–output relationship.

Figure 9 shows the posterior distributions of the differences in the standard deviations of persistent and transient inefficiencies between the two models. These distributions offer insight into whether the variability in inefficiency is significantly affected by the choice of functional form. In Figure 9a (left panel), the mean difference in the standard deviation of persistent inefficiencies is relatively small at 0.011. The 95 % credible interval ranges from −0.045 to 0.068, which includes zero. This indicates that there is no significant difference in the dispersion of persistent inefficiency estimates across the two models. In contrast, Figure 9b (right panel) presents a strikingly different picture for transient inefficiencies. The posterior mean difference in standard deviations is 0.219, with a 95 % credible interval of [0.094, 0.314], which lies entirely above zero. This clearly indicates that the two models yield significantly different levels of variability in the transient inefficiency component. The

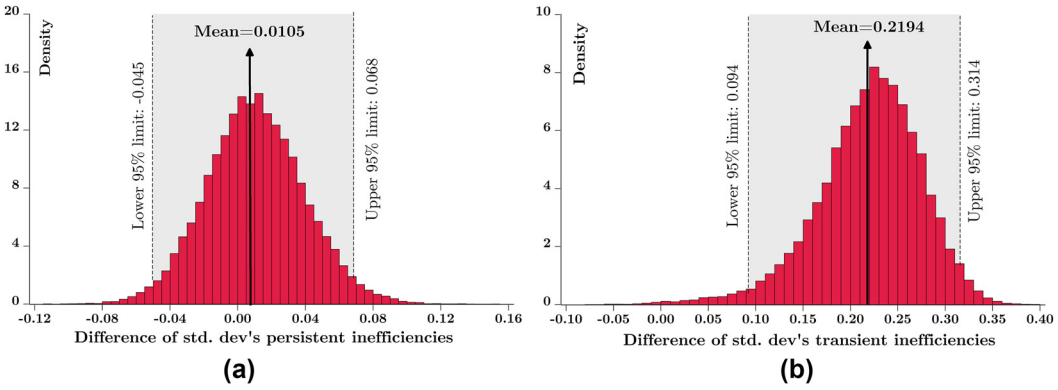

Figure 9: Histograms of differences in the std. dev's of inefficiencies.

Table 1: Comparison of neural network SFA and conventional SFA models.

Model	Marginal log likelihood	Prior probability	Posterior probability	DIC
Neural network SFA	−9.442	0.5	1	1,270.900
SFA	−38.492	0.5	0	1809.200

greater dispersion observed under the Cobb–Douglas model suggests that it may inflate the heterogeneity in time-varying inefficiency compared to the more flexible neural network specification.

Finally, Table 1 compares the marginal log likelihoods, prior probabilities, posterior probabilities, and DIC values for the neural network SFA and conventional SFA models. The neural network SFA model achieves a posterior probability of 1 and a lower DIC, indicating a clearly better fit to the data compared to the conventional specification. Combined with the results in Figures 8 and 9, this provides strong empirical evidence that the Cobb–Douglas model not only underperforms in terms of overall fit but also systematically inflates transient inefficiency. This highlights the importance of using more flexible functional forms, such as neural networks, when estimating production frontiers and inefficiency components.

5 Concluding Remarks

This study introduces a Bayesian neural network SFA model designed to capture non-linear relationships between inputs and outputs while estimating both persistent and transient inefficiencies. This is achieved by specifying a generalized true random effects model in which the deterministic part of the production frontier is modeled using a neural network. Rather than relying on pre-defined functional forms such as the Cobb–Douglas, the neural network learns the input–output relationship from the data by expressing the frontier as a weighted sum of non-linear transformations of the input variables. This flexible structure enables the model to approximate complex production relationships more accurately, thereby improving the reliability of the inefficiency estimates. For comparison, we also consider a generalized true random effects SFA model with the conventional linear Cobb–Douglas functional form. Both models are estimated within a Bayesian framework, and the comparison focuses on selected moments of the inefficiency distributions as well as overall model fit. The models are applied to panel data from Chilean hydroelectric power plants.

The empirical findings from the neural network SFA model reveal complex non-linear relationships between all inputs and output, patterns that are missed by the constant elasticities implied by the Cobb–Douglas specification. Both persistent and, more notably, transient inefficiency scores are inflated under the Cobb–Douglas SFA model compared to the neural network alternative. While the differences in the mean and standard deviation of persistent inefficiencies are not significant, the differences for transient

inefficiencies are both substantial and significant. Moreover, model comparison based on Bayes factors and DIC strongly favors the neural network SFA model, indicating that it provides a better fit to the data.

These findings have important implications for practitioners aiming to assess and improve firm performance. The choice of functional form in stochastic frontier models can substantially influence the estimated levels of inefficiency, particularly transient inefficiency, which reflects short-term performance fluctuations. Relying on a restrictive specification like the Cobb-Douglas may lead to inflated inefficiency estimates, potentially resulting in misguided managerial responses. In contrast, the neural network SFA model, by accommodating non-linear input-output relationships, offers more reliable inefficiency estimates. Moreover, the superior model fit indicated by the Bayes factor and DIC comparison suggests that flexible, data-driven approaches better capture the underlying production technology. Practitioners are therefore encouraged to adopt more adaptable modeling frameworks, such as neural networks, when evaluating efficiency, especially in complex or data-rich environments.

From a policy perspective, the finding that conventional models inflate transient inefficiency has important implications. Transient inefficiency often reflects short-term performance variability that can result from temporary shocks, such as market volatility, or regulatory adjustments. If policymakers rely on estimates derived from restrictive functional forms, they may overstate short-term inefficiency and misinterpret temporary deviations as persistent performance gaps. This could lead to misguided interventions or resource allocations. By adopting more flexible approaches, such as the proposed neural network SFA, policymakers can obtain a clearer picture of firms' true performance dynamics and design more targeted and proportionate policy measures.

Despite its promising results, this study has some limitations that should be acknowledged. The analysis is based on a specific neural network structure with a fixed activation function and number of hidden neurons, which may influence the model's flexibility and performance. Moreover, the empirical application focuses on a single dataset from the energy sector, and thus the generalizability of the findings to other industries remains to be explored. These limitations also point toward meaningful directions for future research.

To further enhance the analysis, future research could explore two potential extensions. First, while the tanh activation function is widely used, alternative activation functions, such as the sigmoid and Rectified Linear Unit, could be tested to determine if they provide better model performance in capturing complex relationships. Second, within a Bayesian framework, the number of neurons in the neural network could be empirically determined using shrinkage priors, as suggested by Hauzenberger et al. (2025).

Data availability: The data and code to reproduce the results are provided as Supplementary Material.

Appendix

Table A.1: Neural network SFA estimates.

Variable	Mean	Std	95 % credible interval	GD	MCSE
$W1_{1,1}$	−18.610	7.414	[−33.143, −4.646]	0.189	0.729
$W1_{2,1}$	1.054	0.606	[0.172, 2.400]	0.340	0.027
$W1_{3,1}$	7.781	1.698	[4.687, 11.419]	−0.352	0.128
$W1_{4,1}$	−5.696	1.340	[−8.475, −3.286]	−0.387	0.100
$W1_{5,1}$	−0.224	0.097	[−0.429, −0.049]	−0.224	0.002
$W1_{1,2}$	6.924	0.769	[5.734, 8.140]	0.485	0.042
$W1_{2,2}$	0.133	0.028	[0.091, 0.196]	0.133	0.003
$W1_{3,2}$	0.056	0.014	[0.032, 0.087]	0.056	0.001
$W1_{4,2}$	0.016	0.007	[0.003, 0.032]	0.016	0.000
$W1_{5,2}$	0.003	0.003	[−0.002, 0.008]	0.003	0.000
$W1_{1,3}$	8.923	7.758	[−6.771, 23.670]	0.367	0.388
$W1_{2,3}$	3.542	1.603	[0.857, 7.052]	0.079	0.802

Table A.1: (continued)

Variable	Mean	Std	95 % credible interval	GD	MCSE
$W1_{3,3}$	−15.941	5.050	[−27.025, −7.549]	−0.316	0.316
$W1_{4,3}$	−4.083	1.555	[−7.701, −1.695]	−0.795	0.078
$W1_{5,3}$	0.089	0.338	[−0.688, 0.696]	0.002	0.011
$W1_{1,4}$	−6.430	7.686	[−19.222, 8.720]	1.319	0.384
$W1_{2,4}$	−3.777	3.726	[−14.024, −0.868]	1.349	0.019
$W1_{3,4}$	6.950	3.060	[2.838, 15.655]	0.685	0.015
$W1_{4,4}$	4.398	4.230	[0.896, 15.751]	0.852	0.021
$W1_{5,4}$	−0.817	2.054	[−7.592, 0.459]	0.658	0.010
$W2_1$	−0.461	0.042	[−0.547, −0.382]	0.999	0.001
$W2_2$	6.816	2.116	[3.771, 11.689]	0.406	0.344
$W2_3$	−0.315	0.035	[−0.387, −0.247]	0.003	0.000
$W2_4$	−1.557	0.142	[−1.875, −1.307]	0.003	0.000
$b1_1$	−18.604	7.854	[−34.356, −2.733]	−0.588	0.835
$b1_2$	−6.944	0.753	[−8.191, −5.734]	12.079	0.413
$b1_3$	9.903	7.809	[−5.439, 25.358]	0.103	0.341
$b1_4$	−8.383	8.369	[−23.812, 8.372]	1.446	1.620
$b2$	−2.033	1.863	[−6.442, 0.565]	−0.189	0.323
ω	685.181	674.124	[72.185, 2,574.721]	0.624	4.192
ψ	20.530	5.942	[10.719, 33.703]	−0.534	0.036
ϕ	11.783	3.205	[7.109, 19.564]	−1.624	0.048
τ	38.473	12.335	[21.593, 68.327]	1.643	0.220

Std stands for Standard Deviation, GD for Geweke Diagnostic, and MCSE for Monte Carlo Standard Error.

Table A.2: SFA estimates.

Variable	Mean	Std	95 % credible interval	GD	MCSE
β_1	1.341	0.208	[0.938, 1.753]	−0.134	0.001
β_2	0.546	0.029	[0.489, 0.602]	−0.049	0.000
β_3	0.194	0.044	[0.108, 0.279]	−0.558	0.000
β_4	0.250	0.035	[0.182, 0.319]	−0.654	0.000
β_5	0.030	0.014	[0.004, 0.057]	1.067	0.000
ω	587.429	626.925	[54.774, 2,301.905]	−0.907	3.782
ψ	19.402	5.841	[9.829, 32.570]	−1.152	0.031
ϕ	2.923	1.034	[1.819, 5.374]	0.769	0.017
τ	12.600	4.240	[6.246, 22.500]	−0.322	0.043

Std stands for Standard Deviation, GD for Geweke Diagnostic, and MCSE for Monte Carlo Standard Error.

References

Aigner, D., C. A. K. Lovel, and P. Schmidt. 1977. “Formulation and Estimation of Stochastic Frontier Production Function Models.” *Journal of Econometrics* 6 (1): 21—37.

Atkinson, S. E., and J. H. Dorfman. 2009. “Feasible Estimation of Firm-specific Allocative Inefficiency Through Bayesian Numerical Methods.” *Journal of Applied Econometrics* 24 (4): 675—97.

Atkinson, S. E., and C. E. Halabí. 2005. “Economic Efficiency and Productivity Growth in the Post-Privatization Chilean Hydroelectric Industry.” *Journal of Productivity Analysis* 23: 245—73.

Azadeh, A., S. F. Ghaderi, M. Anvari, and M. Saberi. 2007. “Performance Assessment of Electric Power Generations Using an Adaptive Neural Network Algorithm.” *Energy Policy* 35 (6): 3155—66.

Azadeh, A., M. Saberi, and M. Anvari. 2010. “An Integrated Artificial Neural Network Algorithm for Performance Assessment and Optimization of Decision Making Units.” *Expert systems with applications* 37 (8): 5688—97.

Giannakas, K., K. C. Tran, and V. Tzouvelekas. 2003. “On the Choice of Functional Form in Stochastic Frontier Modeling.” *Empirical Economics* 28 (1): 75—100.

Hauzenberger, N., F. Huber, K. Klieber, and M. Marcellino. 2025. “Bayesian Neural Networks for Macroeconomic Analysis.” *Journal of Econometrics* 249: 105843.

Heaton, J. T. 2008. *Introduction to Neural Networks for Java*, 2nd ed., 439. St. Louis, MO: Heaton Research, Inc.

Kerstens, K., C. O’Donnell, and I. Van de Woestyne. 2019. “Metatechnology Frontier and Convexity: A Restatement.” *European Journal of Operational Research* 275 (2): 780—92.

Konstantakis, K. N., P. G. Michaelides, P. Xidonas, A. N. Prelorentzos, and A. Samitas. 2025. “Responsible Artificial Intelligence for Measuring Efficiency: A Neural Production Specification.” *Annals of Operations Research* 354: 399—425.

Kumbhakar, S. C., G. Lien, and J. B. Hardaker. 2014. “Technical Efficiency in Competing Panel Data Models: A Study of Norwegian Grain Farming.” *Journal of Productivity Analysis* 41 (2): 321—37.

Lewis, S. M., and A. E. Raftery. 1997. “Estimating Bayes Factors via Posterior Simulation with the Laplace — Metropolis Estimator.” *Journal of the American Statistical Association* 92 (438): 648—55.

Liao, H., B. Wang, and T. Weyman-Jones. 2007. “Neural Network Based Models for Efficiency Frontier Analysis: An Application to East Asian Economies’ Growth Decomposition.” *Global Economic Review* 36 (4): 361—84.

Meeusen, W., and J. van den Broeck. 1977. “Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error.” *International Economic Review* 18 (2): 435—44.

Pendharkar, P. C. 2023. “A Radial Basis Function Neural Network for Stochastic Frontier Analyses of General Multivariate Production and Cost Functions.” *Neural Processing Letters* 55 (5): 6247—68.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. 2002. “Bayesian Measures of Model Complexity and Fit.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (4): 583—639.

Tsionas, E. G., and S. C. Kumbhakar. 2014. “Firm Heterogeneity, Persistent and Transient Technical Inefficiency: A Generalized True random-effects Model.” *Journal of Applied Econometrics* 29 (1): 110—32.

Tsionas, E. G., C. F. Parmeter, and V. Zelenyuk. 2023. “Bayesian Artificial Neural Networks for Frontier Efficiency Analysis.” *Journal of Econometrics* 236 (2): 105491.

Van den Broeck, J., G. Koop, J. Osiewalski, and M. F. J. Steel. 1994. “Stochastic Frontier Models: A Bayesian Perspective.” *Journal of econometrics* 61 (2): 273—303.

Wang, S. 2003. “Adaptive Non-Parametric Efficiency Frontier Analysis: A Neural-Network-Based Model.” *Computers & Operations Research* 30 (2): 279—95.

Xuan, Y. 2018. “Performance Evaluation Research of Ecological Civilization Policy Based on Stochastic Frontier Analysis and Artificial Neural Networks Model.” *Journal of Law, Policy and Globalization* 77: 139—52.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/snde-2025-0091).