



The impact of Clean Spark Spread expectations on storage hydropower generation

Claudia Condemi¹ · Loretta Mastroeni¹ · Pierluigi Vellucci¹ 

Received: 30 January 2021 / Accepted: 14 September 2021 / Published online: 13 October 2021
© The Author(s) 2021

Abstract

Storage hydropower generation plays a crucial role in the electric power system and energy transition because it is the most widespread power generation with low greenhouse gas emissions and, moreover, it is relatively cheap to ramp up and down. As a result, it provides flexibility to the grid and helps mitigate the short-term production uncertainty that affects most green energy technologies. However, using water in reservoirs represents an opportunity cost, which is related to the evolution of plant production capacity and production profitability. As the latter is related to a wide range of types of variables, in order to incorporate it in a large-scale prediction model it is important to select the variables that impact most on storage hydropower generation. In this paper, we investigate the impact of the variables influencing the choices of price maker producers, and, in particular we study the impact of Clean Spark Spread expectations on storage hydroelectric generation. In this connection, using entropy and machine learning tools, we present a method for embedding this expectations in a model to predict storage hydropower generation, showing that, for some time horizon, expectations on CSS have a greater impact than expectations on power prices. It is shown that, if the right mix of power price and CSS expectations is considered, the prediction error of the model is drastically reduced. This implies that it is important to incorporate CSS expectations into the storage hydropower model.

Keywords Storage hydropower prediction · Clean Spark Spread · Energy market expectations · Entropy methods · Machine learning

JEL Classification C6 · C8 · Q4

✉ Pierluigi Vellucci
pierluigi.vellucci@uniroma3.it

Claudia Condemi
claudia.condemi@uniroma3.it

Loretta Mastroeni
loretta.mastroeni@uniroma3.it

¹ Department of Economics, Roma Tre University, Rome, Italy

1 Introduction

In the last decades the problem of reducing green-house gases emissions has obtained increasing worldwide concern and energy policies have been oriented towards a transformation of the global energy sector from fossil-based to zero-carbon sources, undertaking the so-called *Energy Transition*.

In this scenario, hydropower is of paramount importance as it is the most widespread source of electricity with low GHG emissions. In fact, it generates about 60% of renewable electricity and has *median life-cycle* carbon equivalent intensity of 18.5 gCO₂-eq/kWh (IHA 2018).

Among hydropower plants, those equipped with storage technology play a crucial role for the electric power system and for the energy transition. As known, this technology allows water resources from watercourses or lakes to be stored in reservoirs, and this enables plant management to choose whether and when to release water to produce electricity (Killingtveit 2019). As a result, storage hydropower plants offer flexibility to the grid and help mitigate the short-term production uncertainty that affects most green energy technologies, as shown in Albadi and El-Saadany (2010) and Hirth (2016) for wind and in Komiyama and Fujii (2014) for solar power.

Therefore, it is clear that modelling and predicting hydropower production is of great importance to determine the effects of the energy policy of the countries involved. However, since national production is the sum of generation of several plants, it will be affected by the decisions of a large number of economic players, which makes this modelling troublesome. Besides, national production is planned by geographical sub-areas (market zones) through the articulation of different markets and this particular architecture increases the degree of interaction of the operators' individual choices, making the problem even more complex.

In each market zone, most of the production is planned in spot markets, which are usually auction markets in which bids are regulated according to the merit order criterion. This criterion assumes that power plant operators bid at their marginal costs, which, in case of storage hydropower plants, equals the production opportunity cost (Aasgård et al. 2019). In particular, it depends on the evolution over time of the *production capacity* and the *profitability* of the plant.

As for the production capacity of a storage hydropower plant, it depends on the volume of water in the reservoir (water availability), which in turn is closely related to the volume of water in the lakes and rivers connected to each plant. Water availability has been extensively discussed in the literature. We refer to the following works for more details (Muñoz and Sailor 1998; Cuo et al. 2011; Castillo-Botón et al. 2020; Chen et al. 2019; Ahmad and Hossain 2019; Plucinski et al. 2019).

Concerning profitability, it is defined as the economic convenience of using water at one time instead of another. This problem has been addressed in its various aspects in several works, as, e.g. in Singh and Singal (2017) and Nandalal and Bogardi (2007). In particular, some of these works have pointed out that, when a competitive market involves a number of price-maker producers, the revenue for each producer depends on the bids of all other price-maker producers. In this regard, for instance, Baslis and Bakirtzis (2011) formulate the optimal medium-term scheduling within a unique

stochastic Mixed Integer Linear Programming problem, focusing on the influence of demand variations and competitors' offers on the producer's bidding strategy.

Then, also Steeger and Rebennack (2015) study the bidding problem for multiple price-maker hydropower producers competing in a deregulated, bid-based market. Unlike Baslis and Bakirtzis (2011), their model exploits a Mixed Integer Linear Programming based on discrete functional parameters, highlighting the relevance of price-maker producers operating in the market.

In their work, Birkedal and Bolkesjø (2016) analyze the impact of drivers influencing hydropower scheduling on weekly hydroelectric generation, using a two stage least squares model. This work highlights that hydro balance, inflow and marginal costs of coal-fired power plants are important factors to explain weekly hydropower supply.

Jahns et al. (2020) derive supply curves for hydro-reservoirs in Norway and apply the resulting ones in a multi-region electricity market model, showing how they can be used to perform historical and counterfactual simulations. A key assumption of their model is that an increase in the marginal costs of substituting thermal power plants matches an increase in the water value of reservoirs.

On the other hand, when it comes to large-scale storage hydropower prediction, only a few works have been carried out. In particular, Li et al. (2016) address the problem of annual hydropower forecasting in Japan using Grey models and combining it with Markov chains to improve forecast accuracy.

Monteiro et al. (2014) exploits numerical weather prediction tools (*NWP*) to forecast the hourly aggregate generation in Spain and Portugal. In their work the authors identify a sigmoidal relationship between hydrological power potential and hourly hydroelectric generation, achieving satisfactory results for the next-day forecasts.

Wang et al. (2017) propose the Data Grouping approach based on Grey Modelling (*DGGM*) to forecast the quarterly hydropower production in China. They show that their model performs better than other models such as SARIMA and Grey Model. In particular, they obtain good results for the pre-2011 time series, and quite high errors for the series from 2011 to 2015.

Uzlu et al. (2014) estimate annual hydraulic energy production in Turkey using the Cosine Amplitude Method (*CAM*) to determine the most sensitive factors affecting hydroelectric generation. However, *CAM* is a linear method and, as such, it suffers from the same limitations of commonly used methods (cross-correlation, principal component analysis, cognitive mapping and sensitivity analysis), i.e. it is not able to detect possible nonlinear links among the variables.

Generally speaking, the difficulty of formulating a hydropower model for a large area lies in the ability to select and aggregate the best variables that influence water availability and profitability at a macro-territorial level. This process requires careful simplification because the resulting variables must provide good significance to the problem.

Regarding the economic variables influencing profitability, as argued by Moreno (2009), their consideration in a prediction model implies the introduction of a high level of complexity, due to the large number of different types of potential variables to be used. In fact, the power system is structured into many integrated markets, with different operating rules and purposes for the various steps in the electricity chain.

In this scenario, as argued by Uzlu et al. (2014) and Chen et al. (2016), it is crucial to select the correct predictor variables, because increasing the number of independent variables (size of input vector) can result in over-fitting or over-training the problem, which would reduce the accuracy of the model prediction.

As shown by Condemì et al. (2021b), the main potential drivers of the daily aggregate hydropower generation are the economic variables that embody operators' expectations on short to medium term market trends. In particular, expectations on power prices and on thermal plants' profitability are extremely important. In fact, since it is not possible to purchase water resources, the crucial point in managing hydropower plant resources is to choose the most convenient timing to exploit them. As a consequence, hydropower plant management needs an accurate prediction of future inflow and returns. Obviously, given the uncertainty of these future values, each operator bases her decisions on her own expectations. These expectations are incorporated in the forward prices, which represent the current value that the market attributes to the production of 1 MWh in its delivery period.

The aim of this paper is to study the impact of *Clean Spark Spread* expectations on storage hydroelectric generation (SHG) and to present a method for embedding this expectations in a model to predict monthly storage hydropower generation. In particular, we show that future SHG depends more on the CSS expectations than on those on power prices.

More in detail, we are first interested in detecting the main economic variables that influence storage hydropower generation. Since the main competitors of hydropower producers are thermal producers, these variables correspond to the future revenue expectations of hydropower and thermal producers. To quantify these expectations, we use the corresponding forward values as proxies.

Afterwards, in order to find the best predictors, we exploit an entropy approach to seek the set containing forward on CSS and power prices with the "highest information content" for the prediction of *SHG*. We show that the best set is composed mostly of CSS forward values. Let us remark that we adopt an entropy approach because the relationship among SHG and its drivers is a nonlinear one (Condemì et al. 2021b).

In particular, the approach we follow in this paper is to exploit the creation of a sorted list of conditional entropy computations between different subsets of features and output variables. Then, among them, we pick those with the smallest conditional entropy (for the conditional entropy approach see, e.g. Rastrow et al. 2011; Fischer and Alemi 2020; Wen et al. 2017; Friston et al. 2012).

Finally, we test the prediction ability of these variables in a real problem of monthly hydropower generation in Northern Italy, using Machine Learning models. These models have been widely used in economic and financial analysis, because, compared to statistical/econometric methods, they allow to deal with problems with complex structures (Ghoddusi et al. 2019). The results achieved suggest that the use of *CSS* expectations to predict storage hydropower generation provides competitive results. Moreover, it is clear from the analysis that *CSS* plays a key role in determining hydroelectric generation.

The rest of the paper is structured as follows: in Sect. 2 we analyse the decision-making strategy of the main producers operating in power market and provide proxies for their profitability expectations. Section 3 provides a summary of the methods we

apply to this problem. In Sect. 4 we present the results of our analysis in the real case of the Northern Italy. Section 5 closes the paper with final conclusions and remarks on the research carried out.

2 Problem statement

2.1 Electricity supply chain

The electricity supply chain is composed of all the steps from energy production to its consumption. These include: generation, whole trade, transmission, dispatching, distribution and metering, and retail sale. Among these, transmission, dispatching and distribution sectors are considered to be a natural monopoly, while the other sectors are undergoing liberalization. The centralization of the electric power system is mainly due to the need to keep the grid continuously in balance, i.e. the electricity loaded into the grid, net of losses, must coincide with the electricity consumed.

In this framework, the rules adopted by each country to organize the production of electricity are oriented to minimize the risk of overload or underload of the grid, respecting the principle of free competition. Consequently, in the electricity markets, producers' offers are accepted according to zonal criteria and market rules. In particular, most of the production is organized in the spot electricity markets, which are typically a day-ahead auction markets, regulated according to the merit order criterion (Weron 2014). The merit order model assumes that power plant operators bid at their marginal costs in the electricity spot market, and, since the latter is related to power plant technology, the bids will be homogeneous by type of production.

More specifically, photovoltaic, wind and run of river are technologies that exploit energy resources that cannot be stored and, for this reason, generally offer their production on the markets accepting any price (i.e. they are price-taker operators). Also nuclear electricity producers are price-taker operators, due to their lack of flexibility in operating power plants, which in some cases can also lead to negative price selling.

As for the bids of thermal operators, these are planned according to the operating margin of production, i.e. the difference between the price of the energy produced and the cost of the fuel used to produce it.

In most of the countries,¹ also the amount of environmental taxes should be taken into consideration and subtracted from this value. In particular, in countries that adopt the Emission Trading System (*ETS*), which consists of a cap-and-trade system organized by a regulation authority, each carbon credit certificate authorizes the emission of 1 ton of *CO*₂.

Accordingly, in these countries, gross operating margin of thermal power producers is represented by the Clean Spread, defined as

$$\text{Clean Spread} = \text{electricity price} - \text{fuel price} - \text{carbon credit price}$$

¹ https://carbonpricingdashboard.worldbank.org/map_data.

Based on the fuel exploited by the plant, this value is called Clean Spark Spread (for gas-fired plants) or Clean Dark Spread (for coal-fired plants) and is defined, respectively, as

$$CSS = P_{pw} - P_{gas} - \alpha * P_{CO2}, \quad (1)$$

$$CDS = P_{pw} - P_{coal} - \beta * P_{CO2}, \quad (2)$$

where P_{pw} is the selling power price per MWh , P_{gas} e P_{coal} are, respectively, the prices of gas and coal used to produce 1 MWh . Besides, P_{CO2} is the price of a carbon credit certificate and the parameters α and β represent the number of tons of CO_2 emitted to produce 1 MWh using gas or coal, respectively.

As for hydropower plants, they produce electricity by exploiting the kinetic energy of falling or fast-running water. Among them, storage hydropower plants store water from lakes and rivers in reservoirs. We point out that these water resources can also be brought into the reservoirs by electromechanical lifting with pumping systems.

The bidding strategies of hydro-power producers are based on two criteria. The first, so-called *profitability*, is based on today's production returns compared to future ones. The second criterion is based, instead, on the quantity of the water available in the reservoir, which in turn determines the so-called hydropower *production capacity*, mainly influenced by rainfall and snowpack melting.

2.2 Hydropower economic prediction drivers

To accurately predict storage hydropower generation, a model based on the fundamentals must include both physical variables related to the production capacity of the plants (Ph) and economic variables (Ec) representing the profitability of production and the aggregate power market mechanisms. For this reason, we can express the Storage Hydropower Generation (SHG) as a general nonlinear function of physical and economic prediction variables, as the following

$$SHG = f(Ph; Ec).$$

Concerning profitability, since it is not possible to buy water resources, the crucial point in managing hydropower plant resources is to choose the most convenient timing to exploit them. As a consequence, the hydropower plant management needs an accurate future inflow prediction, and, in addition, has to compare current returns and potential future returns. Obviously, given the uncertainty of these future values, each operator bases her decisions on her own expectations. These expectations are incorporated in the forward prices, which represent the current value that the market attributes to the production of 1 MWh in its delivery period.

Therefore, the current price of a Daily (D), Weekly (W), Monthly (M) or Quarterly (Q) forward corresponds to the actual value of 1 MWh generated in the hours of a specific day, week, calendar month or quarter, respectively. In particular, quarters are fixed as the following: January–March, April–June, July–September, October–December.

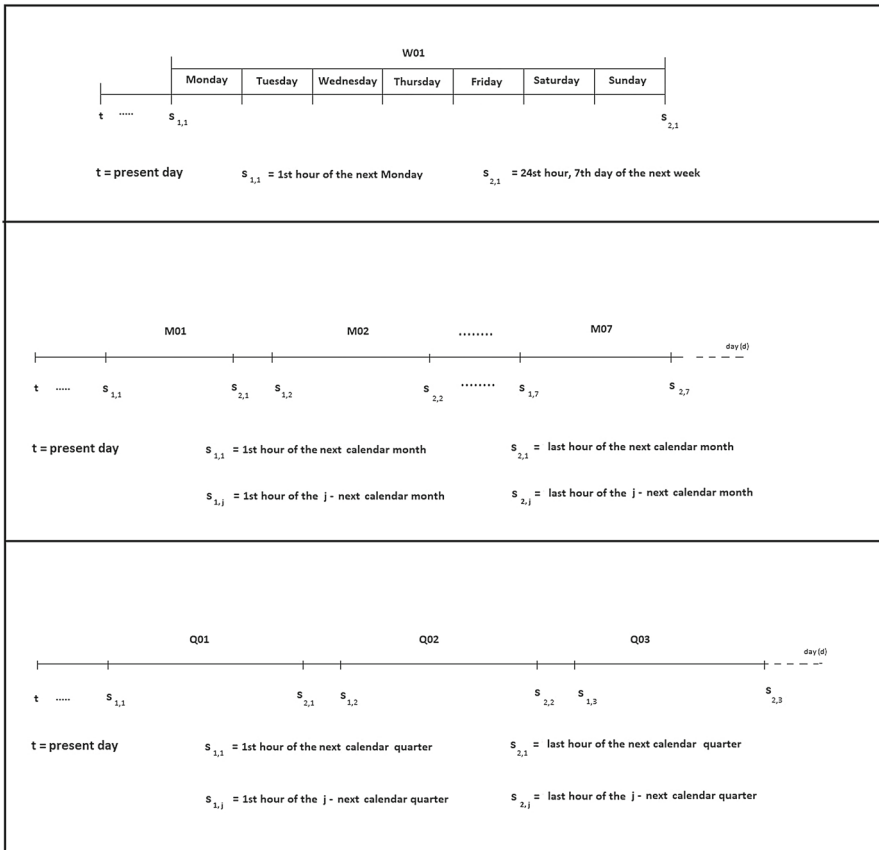


Fig. 1 Forward order per Relative Maturity

Then, for each of these products, the maturity period is a different time interval (s_1, s_2) . For example, as for a monthly forward, s_1 corresponds to the first hour of the first day of the month, and s_2 corresponds to the 24th hour of the last day of the month.

If we consider the time series of forward prices F based on their relative maturity period, we have that $D01$ is the current price of the daily forward with delivery period tomorrow, $M01$ is the current price of the monthly forward with delivery period in the next calendar month, and so on. Generalizing, let today be the day d of the month m of the quarter q . Then we can define DX as the current price of the daily forward with delivery period in the day $d + x$, MX as the current price of the monthly forward with delivery period in the month $m + x$, and finally QX as the current price of the quarterly forward with delivery period in the $m + q$ quarter, respectively (see Fig. 1).

As a result, it is possible to approximate operators' expectations on the gap between the present value of future and current returns by the price spread between the forward FX_{pw} with relative maturity X and daily forward with relative maturity the next day

(D01), as defined in the following

$$\Delta FX_{pw} = FX_{pw} - D01, \quad (3)$$

However, the hydropower generation is influenced by the bids of other price-maker producers, since their bids influence both prices and the amount of demand that be satisfied by hydropower generation. In particular, thermal producers are an important category of price-makers and their bids, as argued above, are formulated according to the Clean Spread. Since there is no forward associated with the Clean Spread, we can quantify the market's expectations on its value using the corresponding forward prices, which in the case of the Clean Spark Spread can be expressed as

$$CSSFX = FX_{pw} - FX_{gas} - \alpha * FXCO2, \quad (4)$$

where FX_{gas} represents the gas forward price in €/MWh with relative maturity X and $FXCO2$ represents the CO₂ forward price with relative maturity X . Let us remark that the parameter α represents the tons of CO₂ emitted to produce 1 MWh. Differently from the hydropower case, since gas can be purchased on the market, it is possible to refer to the Clean Spread in absolute terms. In Table 7 of Appendix A the notations used in the paper are listed.

The aim of this paper is to study the impact of Clean Spark Spread expectations on the aggregate monthly hydroelectric generation and to provide a method to embed this expectations in a storage hydropower generation model. To this end, we analyse the optimal set of the prediction variables of the storage hydropower generation (SHG) using an entropy approach.

As a first step, since usually the interactions among financial and economic variables do not occur simultaneously (but there is a time delay of the effect of the phenomenon on the other variables), it is very important to evaluate the time-delay among the input and the output variables. In addition, there is often a persistence of the effect of one variable on the other over time, so neglecting the lag may lead to assessing the secondary effects of a phenomenon and not the root cause of the phenomenon itself, compromising the correct interpretation of the results.

In order to identify the time lag, we need to investigate the timing of the market. Typically, spot electricity markets are organized as auction markets. Specifically, they are the *day-ahead* market, where most of the production is organized and the *intraday* market, mainly used to make secondary adjustments.

In the day-ahead market, before a certain closing time on day $t - 1$, agents must submit their bids and offers for the delivery of electricity during each hour of day t (see Fig. 2).

Consequently, it is reasonable to assume that the traders' bids at $t - 1$ are based on the knowledge of financial variables at $t - 2$. For this reason, in our analysis we consider the generation at t as a function of the information available at the end of day $t - 2$, as shown in the following function

$$SHG_t = f(Ph_{t-1}, Ec_{t-2})$$

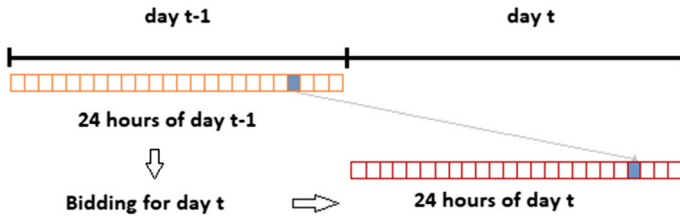


Fig. 2 Bid in day-ahead auction market

3 Methods

In this section we describe the tools employed to identify the best subset of the input variables to predict *SHG*. First, we introduce the basic framework of the entropy analysis of time series, based on the key tools of our approach, i.e. conditional Shannon entropy and transfer entropy. Then, we recall the techniques employed to estimate the entropy measures and, finally, we discuss the role of variable selection in our framework.

3.1 Information measures

The entropy of a random variable is the average level of *information* or *uncertainty* inherent in the variable’s possible outcomes. Formally, the concept of (information) entropy $H(X)$ of a random variable X with a probability mass function $p(x)$ is defined by

$$H(X) = - \sum_x p(x) \log_2 p(x).$$

(see, e.g. Cover and Thomas 2012).

Since we use logarithms to base 2, entropy will be measured in bits. Entropy is a measure of the average uncertainty in the random variable X and corresponds to the number of bits required on average to describe the random variable (Cover and Thomas 2012).

It is possible to define the *conditional entropy* $H(Y|X)$, which is the entropy of a random variable Y , conditioned to the knowledge of another random variable X . Let $p(x, y)$ be the joint probability of these variables, X and Y , occurring together. Hence, conditional entropy $H(Y|X)$ is defined as

$$H(Y|X) = - \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x, y) \log_2 \frac{p(x, y)}{p(x)} \tag{5}$$

where \mathcal{X} and \mathcal{Y} denote the support sets of X and Y .

The conditional entropy (CE) of a random variable Y given another random variable X is zero if and only if Y is a function of X . Hence we can estimate Y from X with zero probability of error if and only if $H(Y|X) = 0$. Extending this argument, we

expect to be able to estimate Y with a low probability of error only if the conditional entropy $H(Y|X)$ is small. On the other hand, it results

$$H(Y|X) \leq H(Y) \quad (6)$$

where the equality holds if and only if X and Y are independent random variables. Accordingly, it is possible to define the *normalized measure of conditional entropy* as the following ratio:

$$r := \frac{H(Y|X)}{H(Y)}. \quad (7)$$

The value of ratio r ranges from 0 to 1. If r is nearer to 1, then we expect that the error made in estimating Y , given X , is high. On the contrary, if r is near to 0, we expect to estimate Y with a low error probability.

In the same way it is possible to define the joint *Shannon* entropy (in bits) of X and Y , as

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x, y) \log_2 [P(x, y)].$$

For more than two random variables X_1, \dots, X_n this expands to

$$H(X_1, \dots, X_n) = - \sum_{x_1 \in \mathcal{X}_1} \dots \sum_{x_n \in \mathcal{X}_n} P(x_1, \dots, x_n) \log_2 [P(x_1, \dots, x_n)], \quad (8)$$

where \mathcal{X}_i denotes the support set of X_i , $\forall i = 1, \dots, n$,

Equation (5) and definitions (6) and (7) can be easily extended to the multivariate case $H(Y|X_1, \dots, X_n)$ by replacing X with random variables X_1, \dots, X_n . Thanks to the multivariate extension of joint entropy (8), $H(Y|X_1, \dots, X_n)$ can also be expressed as the difference between the joint entropy of all variables and the joint entropy of the variables upon which we want to condition:

$$H(Y|X_1, \dots, X_n) = H(X_1, \dots, X_n, Y) - H(X_1, \dots, X_n). \quad (9)$$

The conditional entropy estimated in this paper relies on Eq. (9).

Another tool we exploit in this paper is *Transfer Entropy* (TE).

Brought from information theory, transfer entropy from one random process X to another random process Y is a nonparametric statistic that describes the degree to which X reduces the uncertainty about the future value of Y knowing the past values of X given past values of Y . It allows to detect the direction of the information flow among the time series under study and has the advantage to provide asymmetric interactions, (He and Shang 2017).

In order to define transfer entropy, we assume that the underlying processes evolve over time according to a Markov process (Schreiber 2000). We also combine the

Shannon entropy with the Kullback-Leibler divergence (Kullback and Leibler 1951).² Let us denote by X and Y two sources that emit N symbols with an a-priori joint probability $p(x_i, y_j) := p_{ij}$ and marginal probability $p(x_i) := p_i, p(y_j) := p_j$, whose dynamical structures correspond to stationary Markov processes of order k (process X) and l (process Y). The Markov property implies that the probability to observe X at time $t + 1$ in state i conditional on the k previous observations is $p(i_{t+1}|i_t, \dots, i_{t-k+1}) = p(i_{t+1}|i_t, \dots, i_{t-k})$.

Let $i_t^{(k)} = (i_t, \dots, i_{t-k+1})$ and $j_t^{(l)} = (j_t, \dots, j_{t-l+1})$. Information flow from source Y to source X is measured by quantifying the deviation from the generalized Markov property $p(i_{t+1}|i_t^{(k)}) = p(i_{t+1}|i_t^{(k)}, j_t^{(l)})$ relying on the Kullback-Leibler divergence (Schreiber 2000).

Then, the *Shannon transfer entropy* is given by

$$TE_{Y \rightarrow X}(k, l) = \sum_{i,j} p(i_{t+1}, i_t^{(k)}, j_t^{(l)}) \log_2 \left(\frac{p(i_{t+1}|i_t^{(k)}, j_t^{(l)})}{p(i_{t+1}|i_t^{(k)})} \right), \tag{10}$$

where $TE_{Y \rightarrow X}$ measures the information flow from Y to X ($TE_{X \rightarrow Y}$ as a measure for the information flow from X to Y can be derived analogously).

Transfer entropy is affected from the noise which is present in time series and the noise can lead to misleading results that can be avoided by estimating the *effective transfer entropy* (ETE) (He and Shang 2017). ETE is obtained from the original TE minus the random transfer entropy (RTE). The calculation of ETE is based on the shuffling procedure which is necessary to derive RTE (Behrendt et al. 2019; Dimpfl and Peter 2018; He and Shang 2017; Benedetto et al. 2020) and it is given by:

$$ETE_{Y \rightarrow X}(k, l) = TE_{Y \rightarrow X}(k, l) - RTE_{Y \rightarrow X}(k, l), \tag{11}$$

where RTE is given by:

$$RTE_{Y \rightarrow X} = \frac{1}{N} \sum_{i=1}^N TE_{Y_{shuffled} \rightarrow X}.$$

Data shuffling consists in i.i.d. random draws from the Y time series that are used to generate another time series, i.e. the shuffled series. This procedure eliminates the dependency between Y and X as well as the dependency within Y observations (Behrendt et al. 2019). The shuffling of the series is repeated N times and RTE is obtained from the sample mean of TE where Y is the shuffled sequence. RTE is subtracted to the original TE to obtain ETE estimate as in Eq. (11).

² If we consider two discrete distributions, P and Q , the *Kullback-Leibler divergence* from Q to P is defined as:

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log_2 \left(\frac{P(i)}{Q(i)} \right)$$

3.2 Entropy estimation

By considering Eq. (9), we observe that the multivariate conditional entropy computation requires to determine two entropy terms, namely $H(X_1, \dots, X_n, Y)$ and $H(X_1, \dots, X_n)$. An efficient estimate of entropy is therefore essential to calculate $H(Y|X_1, \dots, X_n)$.

Entropy estimation has gained much interest over the last decades (Meyer 2008) and most approaches focus on reducing the bias inherent to entropy estimation. The methods developed in Meyer (2009) focus on the fastest and most used entropy estimators. We exploit some of these estimators in the case study that we analyze in Sect. 4. Namely they are the empirical estimator and the Miller–Madow bias correction estimator.

Let's now define them. The empirical estimator is the entropy of the empirical distribution:

$$\hat{E}^{emp}(X) = - \sum_{x \in \mathcal{X}} \frac{\#(x)}{m} \log \frac{\#(x)}{m},$$

where $\#(x)$ is the number of data points having value x and m is the number of samples. It can be shown that entropy estimators are biased downwards, and the asymptotic bias is $-\frac{|\mathcal{X}|-1}{2m}$ and depends on the number of bins $|\mathcal{X}|$ (Meyer 2008; Paninski 2003).

As for the Miller–Madow correction estimator, it is defined as the empirical entropy corrected for the asymptotic bias, as in the following

$$\hat{E}^{mm}(X) = \hat{E}^{emp}(X) + \frac{|\mathcal{X}| - 1}{2m}, \quad (12)$$

where $|\mathcal{X}|$ is the number of bins with nonzero probability. This correction, while adding no computational cost, reduces the bias without changing variance. As a result, the Miller–Madow estimator is often preferred to the empirical entropy estimator which proves to be naive.

These estimators, as many others, have been designed for discrete variables. If the random variable X is continuous and taking real values in $[a, b]$, then we have to partition this interval into $|\mathcal{X}|$ sub-intervals in order to employ a discrete entropy estimator. In this paper, following the approach by Meyer (2008), we adopted the equal frequency quantization algorithm. According to this algorithm, the $|\mathcal{X}|$ sub-intervals are such that each of them has the same number of data points, i.e. $m/|\mathcal{X}|$ (Dougherty et al. 1995; Liu et al. 2002; Yang and Webb 2009). The choice $|\mathcal{X}| = \sqrt{m}$ has been proved to be a fair trade-off between bias and variance (Meyer 2008).

Let's turn now to TE. Since time series data are continuous and TE is a discrete measure, original data must be discretized, using symbolic encoding, to estimate the joint probabilities in (10)—for further details see, e.g. Behrendt et al. (2019).

The estimation of the joint probabilities in TE computation is challenging. One can refer, e.g. to Lee et al. (2012) and Behrendt et al. (2019).

A way to obtain the PDFs in Eq. (10) is to allocate data points to fixed, equally-spaced bins. Let us denote the bounds specified for the n bins by q_1, q_2, \dots, q_n , where $q_1 < q_2 < \dots < q_n$, and consider a time series denoted by $X = \{x_t\}$. From a mathematical point of view, we define a function \mathcal{Q} (called *quantizer*) $\mathcal{Q} : x_t \mapsto s_t$ such that

$$s_t = \begin{cases} 1 & x_t < q_1 \\ i & x_t \in [q_{i-1}, q_i) \quad \text{for } i = 2, \dots, n. \\ n & x_t \geq q_n \end{cases} \quad (13)$$

The allocation of data points to equally-spaced bins is less time consuming than other methods to estimate TE as the Nearest Neighbours method but has the drawback of detecting more false positives than the latter (Assis and de Assis 2018). In this paper we employ a $q = 3$ -quantile binning, partitioning the data into three bins through the 5% and 95% empirical quantiles of the data distribution as suggested by Behrendt et al. (2019); Dimpfl and Peter (2018).

Table 2 in Sect. 4.2 provides descriptive statistics for our dataset. All the series (except the SHGN, which, however, is compared with all the others) exhibit an excess kurtosis. Therefore, it seems reasonable to investigate the information contained in the tails of our distributions via TE according to the aforementioned discretization into three bins. This is an established practice in the literature (Benedetto et al. 2020; Behrendt and Schmidt 2020; Behrendt and Prange 2021). Moreover, Behrendt and Prange (2021), for a number of observations comparable to ours, argue that a partitioning into more bins would require more data.

Moreover, we still performed the analysis of TE between SHGN and the other series, by increasing the number of quantiles incrementally from 1 to 10, as done by Park et al. (2021); see Appendix B for more details.

3.3 Variable selection

The variable selection problem is often defined as the selection of a subset of variables based on statistical estimates of its performance and can be considered as a particular form of model selection (Reunanen 2003). It is an important step in building an automatic predictor (e.g., the best one). The accuracy of the prediction can be improved by excluding irrelevant variables, and, at the same time, variable selection increases the intelligibility of a model, even though we cannot ignore the fact that by eliminating a variable, we lose its information.

Let $X = (X_S, X_R)$ be composed of two subsets of variables, X_S , standing for the selected variables, and X_R , the remaining or eliminated variables (Meyer 2008).

By definition of *conditional mutual information*³ (Meyer 2008; Cover and Thomas 2012), we have,

$$H(Y | X) = E(Y | (X_S, X_R)) = E(Y | X_S) - I(X_R; Y | X_S),$$

where $I(X_R; Y | X_S)$ denotes the conditional mutual information of the random variable X_R and Y given X_S . If

$$I(X_R; Y | X_S) > 0,$$

i.e. if X_R possesses some information on Y given X_S , then eliminating X_R increases the uncertainty on the output variable. In other words:

$$E(Y | (X_S, X_R)) \leq E(Y | X_S)$$

However, eliminating information could increase noise but improves the reliability (less variance) of the estimation — see, e.g. the bias-variance trade-off as discussed by Meyer (2008).

The approach we follow in this paper is to perform a feature selection by creating a sorted list of conditional entropy computations between the different subsets of features and the output variable and then picking those with the smallest conditional entropy. Minimizing the conditional entropy is a task that can be found in a plenty of applications (Rastrow et al. 2011; Fischer and Alemi 2020; Wen et al. 2017; Friston et al. 2012).

Let us formalize this process. Clearly, for n input variables, the number of possible subsets is 2^n . This step entails finding the best subset of variables in the power set $2^{\mathcal{S}}$ where \mathcal{S} denotes the set of random variables X_1, \dots, X_n . Hence, it is an example of combinatorial optimization problem (Kohavi and John 1997; Meyer 2008). More formally the problem is, given n input variables X_1, \dots, X_n , find the subset $S_0^{\max} \subset \mathcal{S}$ which minimizes the conditional entropy

$$S_0^{\max} = \arg \min_{S_0 \in 2^{\mathcal{S}}} E(Y | S_0).$$

4 Experiments and results

In this section, we analyze the case study related to storage hydropower generation in Northern Italy (*SHGN*).

³ For discrete random variables X , Y , and Z the *conditional mutual information* $I(X; Y|Z)$ is defined as follows

$$I(X; Y|Z) = \sum_{z \in \mathcal{Z}} \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p_{X,Y,Z}(x, y, z) \log \frac{p_Z(z) p_{X,Y,Z}(x, y, z)}{p_{X,Z}(x, z) p_{Y,Z}(y, z)}.$$

The analysis is structured in three steps. In the first, we estimate the transfer entropy between the storage hydropower generation series and the proxy time series of CSS expectations, identified as argued in Sect. 2. Here we have employed the R package *Rtransferentropy* for the TE computation, which heavily relies on the method reported in Behrendt et al. (2019); Dimpfl and Peter (2018).

In the second step, we estimate the conditional entropy between *SHGN* and several sets of variables, to identify the best set of economic variables for predicting *SHGN*. To this purpose, we initially analyse the set of variables directly correlated with hydroelectric operators' revenues, i.e. proxies of future power prices. Then, we exploit the results of the first step as a guideline to obtain the sub-set of variables with the highest information content.

CE has been estimated by means of the R package *infotheo*, based on the Meyer's work (Meyer 2008, 2009).

In the final step, we compare the predictive performance of different *SHGN* models based on the sub-sets identified in the previous step and a machine learning (ML) approach.

4.1 Northern Italy power market

The Italian electricity transmission grid is partitioned into virtual and geographical zones. Virtual zones correspond to points of interconnection with foreign countries, called *foreign virtual zone*, and to limited production poles, called *national virtual zone*. Instead, *geographical zones* represent a portion of the national network relating to a geographical area. In particular, there are 6 geographical zones: Northern Italy, Central Northern Italy, Central Southern Italy, Southern Italy, Sardinia and Sicily (see Fig. 3).

Among them, Northern Italy has the highest share of hydroelectric generation.⁴ It covers a geographical area of 6392.17 km² (ISTAT 2020), including the Italian Alps, where storage hydropower plants are mainly located (see Fig. 4). Table 1 provides the main information on these plants.

With regard to other types of producers, we do not consider the production dynamics related to price-taker producers since they do not provide any information on hydropower generation (Condemi et al. 2021b). Instead, variables affecting the competitiveness of thermal power plants play an important role. In particular, in Northern Italy, gas-fired power plants work as base load plants, whose gross operating margin is represented by the Clean Spark Spread (1). For this reason, we do not include CDS (2) in the real case application.

4.2 Data description

$SHGN_{d_x}$ represents the total amount of electricity (in GWh) generated, in the whole area of study, during the day d_x by storage hydro-power plants.

⁴ This power market zone is the geographic area composed of the following regions: Valle d'Aosta, Piemonte, Liguria, Trentino, Veneto, Friuli Venezia Giulia and Emilia Romagna.

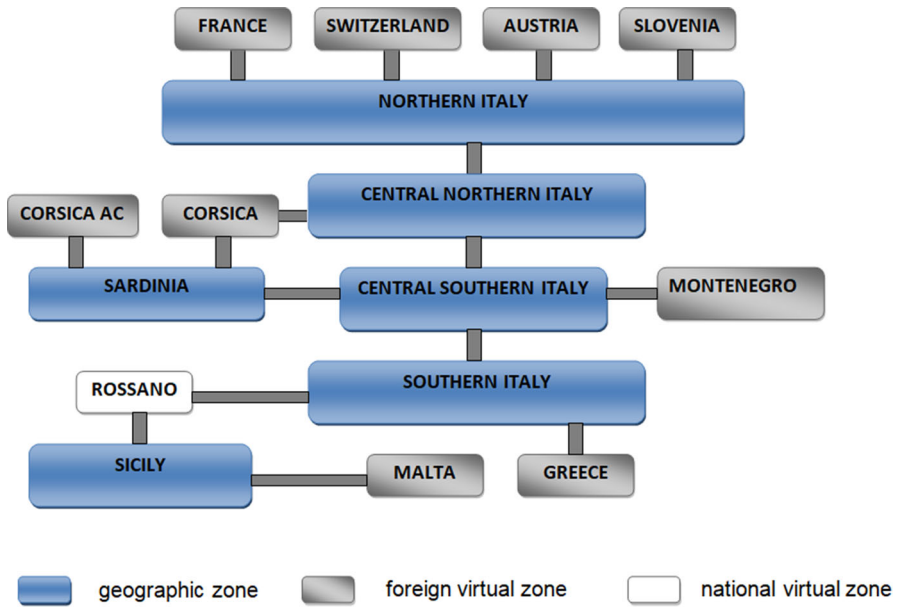


Fig. 3 Topology of interconnection among the zones

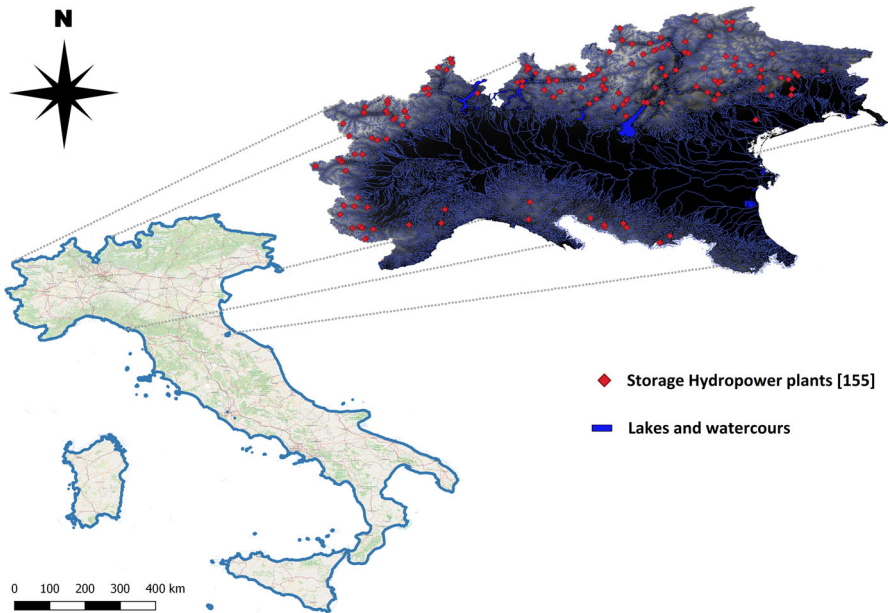


Fig. 4 Distribution of the storage hydropower plants and hydrography of Northern Italy

Table 1 Storage Hydropower plants characteristics in Northern Italy

Size	N. plants	Nominal power β				
		Total (MW)	Alps (%)	Other (%)	Mean (MW)	Std. Dev. (MW)
Large plants ($\beta > 100$ MW)	30	8628.31	96.74	3.26	287.61	243.09
Medium plants ($10 < \beta \leq 100$ MW)	110	3638.08	96.18	3.82	33.07	21.00
Small plants ($\beta \leq 10$ MW)	15	138.48	54.59	45.41	9.23	0.76
Total	155	12404.87	96.11	3.89	80.03	148.80

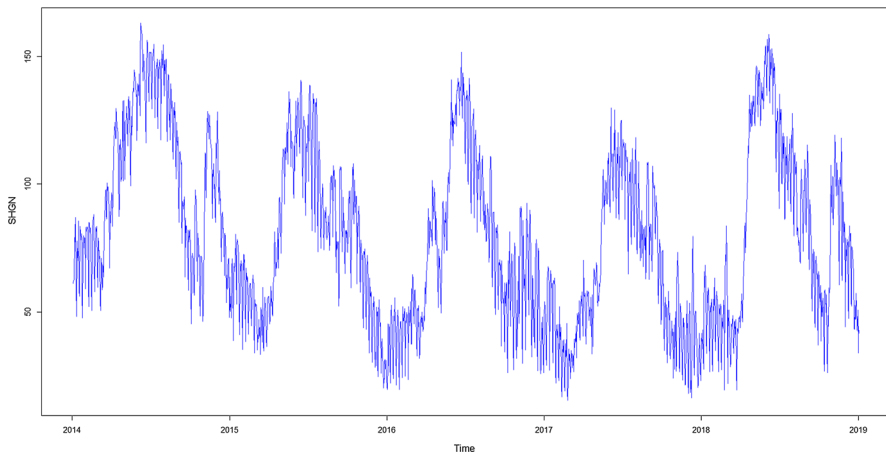


Fig. 5 Hydroelectricity (in GWh) generated by storage plants in Northern Italy, from 04/01/2014 at 03/01/2019

In our work, we examine the time series of daily electricity generated by storage hydropower plants in Northern Italy ($SHGN_t$), collected from 04/01/2014 to 03/01/2019 ($t_x = d_x$). $SHGN_{d_x}$ (Fig. 5) represents the total amount of electricity (in GWh) generated, in the whole area of study, during the day d_x by storage hydro-power plants (data source: TERNA). In accordance with the time lag assessment performed in Sect. 2, time series of proxies, ΔFX_t (3) and $CSSF X_t$ (4), have been synchronised so that $t_x = d_x - 2$. Consequently, these series refer to the days (d) from 02/01/2014 to 01/01/2019. In particular, our elaboration is based on EEX (2019) data.

As regard the maturity period, we compute the values of daily proxies for the following maturity periods (X): $D01, M01, M02, M03, M04, M05, M06, M07, Q01, Q02, Q03$. Moreover, since there are no one-week forwards written on Gas, the proxy referring to maturity $W01$ has been computed only for $\Delta W01$ (power price proxy). In particular, to compute the value of such proxies, we use the average of the

market closing prices referring to Forwards written on the PUN⁵ and Italian PSV⁶ Natural Gas price. Conversely, we estimate the parameter α year by year based on ISPRA (2018) data.

As regards the gaps in the time series caused by market closures, we assume that producers base their decisions on the latest available data.

Table 2 contains the main time series statistics used in this paper. In particular, we apply the Dickey-Fuller univariate tests, ADF and PP, based on Banerjee's et al. tables and on J.G. McKinnons' numerical distribution functions (Banerjee et al. 1993), and we adopt the *median absolute deviation* (mad), as index of variability. The column "Kurtosis" of Table 2 shows that the distribution of our data are all leptokurtic, except SHGN.

4.3 Transfer entropy analysis

In this section we investigate the influence of the Clean Spark Spread forward values on *SHGN*, using the proxies defined in Sect. 4.2. To this aim, for each of the eleven *CSSF* time series, we estimate the effective transfer entropy, defined in Eq. (11), from *SHGN* (Y) to *CSSF* (X), $ETE_{Y \rightarrow X}$, and in the opposite direction, $ETE_{X \rightarrow Y}$.

Furthermore, to establish the dominant direction in the relationship between X and Y , we use the following criterion: if $ETE_{Y \rightarrow X}$ and $ETE_{X \rightarrow Y}$ have similar values or both have values approximately equal to zero, we define the dominant direction between X and Y as doubtful. Instead, if $ETE_{Y \rightarrow X}$ and $ETE_{X \rightarrow Y}$ have strictly distinct values for all iterations, we define the dominant relationship as that relating to the ETE with the highest values.

Let us consider Figs. 6 and 7, and Table 3. Using the denomination introduced in Benedetto et al. (2020), figures and table contain, respectively, static and dynamic transfer entropy analysis. The static transfer entropy is a number, and it is estimated over the entire sample size. The dynamic transfer entropy, instead, is calculated with a growing window approach, always starting from the first observation and increasing window size.

For instance, concerning *CSSM01* (see Fig. 6), it is evident that series $ETE_{Y \rightarrow X}$ has always greater values than the series $ETE_{X \rightarrow Y}$. Consequently, regarding the maturity *M01*, we define the direction from *CSSM01* to *SHGN* as the dominant one.

Table 3 shows the results of the Transfer Entropy analysis between *SHGN* and *CSSF* for their respective maturities. As the results clearly indicate, the information from *CSSF* to *SHGN* is the most relevant.

In fact, there is a clear dominance for the following maturities: *D01*, *M01*, *M03*, *M04*, *M06*, *M07*, *Q02*, *Q03*. Therefore, the knowledge of Clean Spark Spread expectations for these maturities is important to predict storage hydropower generation and thus including them in a *SHGN* model will improve its performance.

⁵ "Prezzo Unico Nazionale" (PUN) is the average of day-ahead market zonal prices, weighted for the related Volumes, of all transactions executed during a market session.

⁶ PSV, "Punto di Scambio Virtuale" is the Italian Virtual Trading Point organized and managed by Snam Rete Gas.

Table 2 Descriptive statistics

	Mean	Mad	Min	Max	ADF Statistic (<i>p</i> value)	PP test		Kurtosis
						Z (alpha)	(<i>p</i> value)	
SHGN	80.15	39.25925	15.32	163.11	-7.6241 (<i>p</i> < 0.01)	-76.937	(<i>p</i> < 0.01)	2.159724
ΔW01	0.5901	5.04084	-58.9900	23.0000	-23.5372 (<i>p</i> < 0.01)	-611.55	(<i>p</i> < 0.01)	13.46398
ΔM01	0.5343	5.911747	-68.5489	27.8271	-17.336 (<i>p</i> < 0.01)	-473.95	(<i>p</i> < 0.01)	12.39155
ΔM02	0.8317	7.251079	-70.5457	25.8700	-14.7829 (<i>p</i> < 0.01)	-351.22	(<i>p</i> < 0.01)	10.7191
ΔM03	0.5005	7.917545	-73.4408	20.7450	-13.2848 (<i>p</i> < 0.01)	-272.92	(<i>p</i> < 0.01)	9.473921
ΔM04	0.02035	8.301395	-74.20079	22.04952	-12.0513 (<i>p</i> < 0.01)	-218.95	(<i>p</i> < 0.01)	7.924388
ΔM05	-0.5036	9.00902	-73.0008	21.8958	-11.4272 (<i>p</i> < 0.01)	-195.63	(<i>p</i> < 0.01)	6.593859
ΔM06	-0.7021	9.310727	-74.2169	22.0495	-11.2923 (<i>p</i> < 0.01)	-192.3	(<i>p</i> < 0.01)	6.004318
ΔM07	-1.128	9.933421	-74.017	24.145	-11.4668 (<i>p</i> < 0.01)	-201.82	(<i>p</i> < 0.01)	5.812968
ΔQ01	0.3991	8.037332	-73.5474	23.4087	-13.6083 (<i>p</i> < 0.01)	-290.38	(<i>p</i> < 0.01)	10.1004
ΔQ02	-0.6877	9.048803	-74.7836	21.9983	-11.4188 (<i>p</i> < 0.01)	-196.41	(<i>p</i> < 0.01)	6.127947
ΔQ03	-1.731	9.132816	-82.991	24.145	-12.2295 (<i>p</i> < 0.01)	-234.02	(<i>p</i> < 0.01)	7.177569
CSSD01	28.057	7.255757	7.337	100.424	-30.1182 (<i>p</i> < 0.01)	-764.71	(<i>p</i> < 0.01)	919.7364
CSSM01	28.50	6.470721	16.23	46.95	-30.1719 (<i>p</i> < 0.01)	-764.71	(<i>p</i> < 0.01)	910.2825
CSSM02	28.57	6.158584	16.43	46.88	-30.1631 (<i>p</i> < 0.01)	-759.6	(<i>p</i> < 0.01)	910.3685
CSSM03	28.11	5.095146	17.51	44.30	-30.1571 (<i>p</i> < 0.01)	-759.65	(<i>p</i> < 0.01)	910.3368
CSSM04	27.56	4.4135	16.88	43.97	-30.1733 (<i>p</i> < 0.01)	-759.45	(<i>p</i> < 0.01)	910.3039
CSSM05	27.05	4.300301	17.09	43.69	-30.1872 (<i>p</i> < 0.01)	-759.33	(<i>p</i> < 0.01)	910.316
CSSM06	26.85	4.555832	18.50	40.11	-30.1888 (<i>p</i> < 0.01)	-759.31	(<i>p</i> < 0.01)	910.3264
CSSM07	26.46	4.429759	17.02	37.85	-30.1925 (<i>p</i> < 0.01)	-759.3	(<i>p</i> < 0.01)	910.3575
CSSQ01	28.03	5.42647	17.57	44.05	-30.1652 (<i>p</i> < 0.01)	-759.56	(<i>p</i> < 0.01)	910.3558
CSSQ02	26.87	4.384875	18.26	42.13	-30.1893 (<i>p</i> < 0.01)	-759.32	(<i>p</i> < 0.01)	910.3422
CSSQ03	25.88	4.171153	17.58	37.20	-30.1988 (<i>p</i> < 0.01)	-759.23	(<i>p</i> < 0.01)	910.3794

Fig. 6 ETE between *SHGN* and *CSSM01*

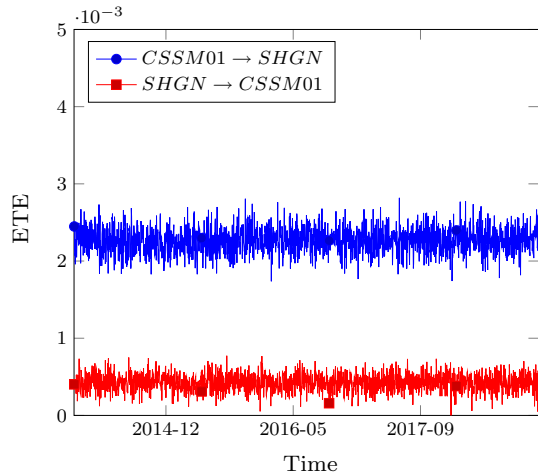
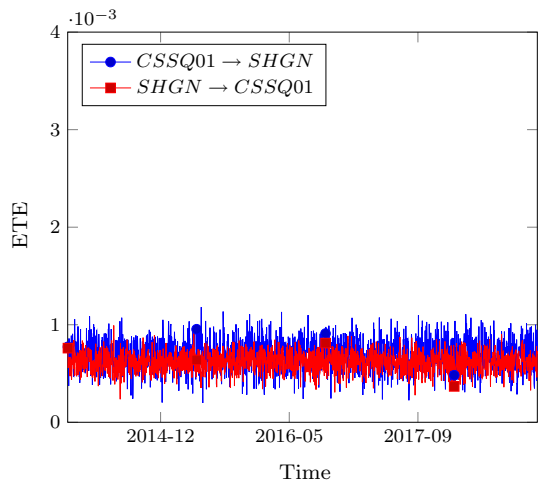


Fig. 7 ETE between *SHGN* and *CSSQ01*



However, there are two cases in which the dominant direction is doubtful. In particular, in the case of *CSSQ01* (see Fig. 7) the values of the two series are very similar, whereas, as regards *CSSM02*, the values of the $ETE_{X \rightarrow Y}$ and $ETE_{Y \rightarrow X}$ series are close to zero, indicating that there may not be a relationship between X and Y . The results enclosed in Table 3 are based on a $q = 3$ -quantile binning. In order to investigate the possible effects of using other values of q and to allow for a robustness check of our results, we performed a TE analysis based on quantiles (see Appendix B for more details).

Therefore, in order to study the relationship relevant to these maturities, further analysis is required. In particular, it is necessary to decide whether to include ΔFX , *CSSF* or no variables. To this aim, a more in-depth analysis is presented in the next section.

Table 3 Effective Transfer Entropy between SHGN (Y) and CSS forward value (X)

Forward value (X)	Average $ETE_{Y \rightarrow X}$	Average $ETE_{X \rightarrow Y}$	Dominant direction
CSSD01	0.0014	0.0082	$X \rightarrow Y$
CSSM01	0.00042	0.0022	$X \rightarrow Y$
CSSM02	0	2.47e-06	Doubtful
CSSM03	0	0.00015	$X \rightarrow Y$
CSSM04	0	0.00050	$X \rightarrow Y$
CSSM05	0.0034	0.00077	$Y \rightarrow X$
CSSM06	0.0007965704	0.0022	$X \rightarrow Y$
CSSM07	0	0.0015	$X \rightarrow Y$
CSSQ01	0.0006186239	0.0006952758	Doubtful
CSSQ02	0	0.00188492	$X \rightarrow Y$
CSSQ03	0	0.001029241	$X \rightarrow Y$

4.4 Conditional entropy analysis

Having assessed, in the previous section, the relationship between the individual $CSSF_X$ and $SHGN$, we are now interested in evaluating the information content of the economic variables as a set of variables (X_i). To this end, we will now estimate the entropy of $SHGN$ (Y) conditional on different sets of variables (X_i), by using the Miller–Madow bias correction estimator, defined by (12).

At first, we separate the products by type, yielding two sets, X_A defined by

$$X_A = (D01, \Delta W01, \Delta M01, \Delta M02, \Delta M03, \Delta M04, \Delta M05, \Delta M06, \Delta M07, \Delta Q01, \Delta Q02, \Delta Q03) \tag{14}$$

and X_B defined by

$$X_B = (CSSD01, \Delta W01, CSSM01, CSSM02, CSSM03, CSSM04, CSSM05, CSSM06, CSSM07, CSSQ01, CSSQ02, CSSQ03)$$

The first is comprised of the power price proxies, ΔFX , whereas X_B is comprised of the $CSSF_X$, except for the $W01$ maturity which, as already mentioned, cannot be represented by CSS prices.

As shown in Fig. 8, the conditional entropy of the set Y given X_A is stable with a mean value of 0.13496, while the set of variables X_B has a lower CE, on average 0.11560. Consequently, compared to set X_A , set X_B contains more information about $SHGN$. However, in the first step of the analysis, we pointed out that there is a marked dominance of the information directionality from $CSSF$ to $SHGN$, with regard to maturities $D01, M01, M03, M04, M06, M07, Q02, Q03$, whereas, as for $M05$, dominance is in the opposite direction. Therefore, by exploiting the transfer entropy analysis, we constructed the set X_C

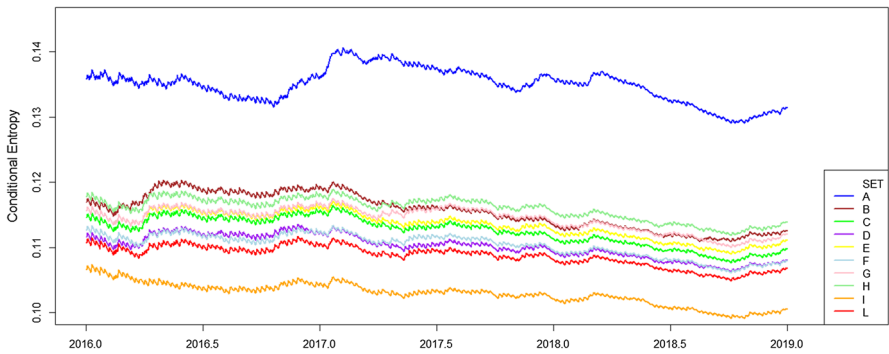


Fig. 8 Conditional Entropy of Y given X_i (2016–2018)

$$X_C = (C SSD01, \Delta W01, C SSM01, C SSM02, C SSM03, C SSM04, \Delta M05, C SSM06, C SSM07, C S S Q01, C S S Q02, C S S Q03)$$

by replacing in the set X_A the proxies related to the maturities for which the dominant direction is from $C S S F$ to $S H G N$. The conditional entropy of $S H G N$ given X_C is stable with a mean value of 0.11252, thus the mixed prices’ set X_C , contains more information about $S H G N$ than the initial sets X_A and X_B .

In addition, as for maturities $Q01$ and $M02$, in the first step of the analysis, a doubtful situation emerged that needs to be clarified in order to identify a better subset of variables. To investigate maturity $M02$, we have estimated the entropy of $S H G N$ conditional on the sets X_D and X_E respectively, where the two sets are defined as follows

$$X_D = (C SSD01, \Delta W01, C SSM01, \Delta M02, C SSM03, C SSM04, \Delta M05, C SSM06, C SSM07, C S S Q01, C S S Q02, C S S Q03)$$

$$X_E = (C SSD01, \Delta W01, C SSM01, C SSM03, C SSM04, \Delta M05, C SSM06, C SSM07, C S S Q01, C S S Q02, C S S Q03)$$

Let us remark that, regarding maturity $M02$, X_D comprises power price proxy, while X_E does not include neither the power price proxy nor the CSS one. The CE corresponding to set X_D and X_E are stable with a mean value of 0.11018 and 0.11345, respectively. Since $H(Y | X_D) < H(Y | X_C) < H(Y | X_E)$, we conclude that, concerning maturity $M02$, it is preferable to include the proxy power $\Delta M02$.

Similarly, as regards $Q01$, we defined the sets X_F and X_G , respectively as

$$X_F = (C SSD01, \Delta W01, C SSM01, C SSM02, C SSM03, C SSM04, \Delta M05, C SSM06, C SSM07, \Delta Q01, C S S Q02, C S S Q03)$$

$$X_G = (C SSD01, \Delta W01, C SSM01, C SSM02, C SSM03, C SSM04, \Delta M05, C SSM06, C SSM07, C S S Q02, C S S Q03)$$

Table 4 Mean Conditional Entropy (2016–2018)

Set	X_A	X_B	X_C	X_D	X_E
Mean $H(Y X)$	0.13500	0.11560	0.11253	0.11018	0.11345
Set	X_F	X_G	X_H	X_I	X_L
Mean $H(Y X)$	0.11039	0.11436	0.11596	0.10291	0.10873

Let us remark that, as for maturity $Q01$, X_F includes power price proxy, while X_G does not include neither the power price proxy nor the CSS one.

As with the case of $M02$, and as shown in Table 4 and in Fig. 8, the highest results are obtained by considering a proxy power price relating to maturity $Q01$. Let us remark that CE plotted in Fig. 8 has been estimated, as done for dynamic TE in Figs. 6 and 7, with a growing window approach, starting from the first observation and increasing window size.

For completeness purposes, we have also compared the CE concerning the sets X_L and X_H , defined as follows:

$$\begin{aligned}
 X_H &= (CSSD01, \Delta W01, CSSM01, CSSM03, CSSM04, \\
 &\quad \Delta M05, CSSM06, CSSM07, CSSQ02, CSSQ03) \\
 X_L &= (CSSD01, \Delta W01, CSSM01, \Delta M02, CSSM03, CSSM04, \\
 &\quad \Delta M05, CSSM06, CSSM07, \Delta Q01, CSSQ02, CSSQ03)
 \end{aligned}
 \tag{15}$$

where as for maturities $M02$ and $Q01$, X_L include power price proxies, while X_H does not include neither power price proxies nor CSS ones. The results in Table 4 show that the best sub-set is X_L with a mean $H(Y|X_L)$ of 0.10873. In particular, given the same number of variables, set X_L has a CE of 19.45%, lower than the set of power price proxy X_A .

It is important to point out that, to provide a rank of importance between power and CSS proxies, we have considered sets containing at most one element per maturity. If we consider, instead, the set X_I composed by all the proxies, defined as follows

$$\begin{aligned}
 X_I &= (CSSD01, D01, \Delta W01, CSSM01, CSSM02, CSSM03, CSSM04, \\
 &\quad CSSM05, CSSM06, CSSM07, CSSQ01, CSSQ02, CSSQ03, \\
 &\quad \Delta M01, \Delta M02, \Delta M03, \Delta M04, \Delta M05, \Delta M06, \Delta M07, \\
 &\quad \Delta Q01, \Delta Q02, \Delta Q03)
 \end{aligned}$$

its CE will result slightly better than the one of X_L , that is about 5.35% less.

However, to achieve this CE reduction we need to double the number of input variables, 23 instead of 12. Nevertheless, in our opinion, the benefits are not sufficient to offset the growth of the complexity following the increase in the number of variables. Therefore, we identify the set X_L as the optimal sub-set for determining $SHGN$.

Finally, if we define the variable y_t as the sum of the *SHGN* from day t to day $t + 30$

$$y_t = \sum_{i=t}^{t+30} SHGN_i \quad (16)$$

the results about $H(y_t|X_i)$ remain unchanged compared to those shown in the case of $H(SHGN|X_i)$.

4.5 Prediction performance

In this section we show the prediction performance of different Storage Hydropower Generation (*SHG*) models based on the economic variables identified in the previous section and on a machine learning approach (Mosavi et al. 2019). As argued in Sect. 2, we defined *SHG* as a nonlinear function of physical (Ph) and economic (Ec) prediction variables related to production capacity and profitability of storage hydropower plants. The output of the trained ML models is the Storage Hydropower Generated in the next 30 days (y_t), as defined in (16).

As regards physical variables, the set Ph_t of input includes daily average values of snow depth (*SW*), rainfall (*Rn*), temperature (*T*) and global solar radiation (*IR*) per hydrological sub-basin (Condemini et al. 2021a)

$$Ph_t = (\mathbf{Rn}_t, \mathbf{SW}_t, \mathbf{T}_t, \mathbf{IR}_t),$$

where t denotes the current day and bold characters denote a vector. In particular, our elaboration is based on the Sistema Nazionale per la protezione dell' Ambiente- SNPA (2019).

Regarding the economic variables, we treated the set X_A (14), including proxies on power price expectations, and the set X_L (15), identified as the best set in Sect. 4.4. We use these two sets alternately to construct the input vector used to train and test the machine learning regressors.

As argued in Sect. 2, we set a time delay of 2 days for economic variables, whereas, concerning the physical one, we set a time lag of 1 day (see Condemini et al. 2021b), defining the variables ya_t and yb_t as, respectively

$$ya_t = f(\mathbf{Ph}_{t-1}, \mathbf{X}_{A;t-2}),$$

and

$$yb_t = f(\mathbf{Ph}_{t-1}, \mathbf{X}_{L;t-2}),$$

In addition, in order to provide a benchmark of the benefit of economic sets for SHGN prediction, we have considered the case in which the input set comprised only the set Ph. Then, we define the following variable

Table 5 Predictive capabilities criterion

MAPE	[0;10]	(10;20]	(20;50]	> 50
Predictive capabilities	High	Good	Reasonable	Weak

$$y_{C_t} = f(\mathbf{Ph}_{t-1}).$$

SVR is used in such a way that the output is always the same (SHGN), whereas, instead, the input matrix varies according to the cases a, b, c above defined.

In all cases, the input database is composed of daily values from 04/01/2014 to 03/01/2019 and has been split in 60% training (1093), 20% validation (365) and 20% testing (365).

We tested different standard machine learning algorithms ⁷ (see Ghoddsi et al. 2019), to compare the two sets of inputs considered. Specifically, we applied nonlinear *Support Vector Regression machine* (SVR), with linear and polynomial kernel (referenced as SVRl and SVRp, respectively), and *Multi-layer Perceptron* (MLP), well-known for its generalization and computational capability (Adnan et al. 2017; Mohd Yassin et al. 2017). In the cases of SVR, the training algorithm used a K-fold cross-validation (with K=5) to select the SVR hyper-parameters. Regarding SVM parameters, the BoxConstraint (C) value is 1 and the Epsilon (ε) value is iqr(Y)/13.49, which is an estimate of a tenth of the standard deviation using the interquartile range of the response variable Y. If iqr(Y) is equal to zero, then the Epsilon value is 0.1. The dataset was splitted as follows: 80% of the data for the training set, 10% for the validation set and 10% for testing. As regard the MLP, the structure used is a two-layer feed-forward network with a sigmoid (17) transfer function in the hidden layer and 10 hidden neurons. We trained this network with the Bayesian Regularization backpropagation function (see Kayri 2016):

$$\tanh = \frac{e^a - e^{-x}}{e^a + e^{-x}} \tag{17}$$

Finally, we evaluated the performance of the corresponding Step-Ahead Prediction Network according to the following metrics

$$\begin{aligned} \text{MAE} &= \frac{1}{N} \sum_{i=1}^N |y_i - \tilde{y}_i| \\ \text{MAPE} &= \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \tilde{y}_i|}{y_i} \end{aligned}$$

According to the level of MAPE, we can define the predictive capabilities of the model as in Table 5 (Wang et al. 2017).

⁷ We specifically use the implementations provided by MATLAB language program (version R2021a with the *Statistics and Machine Learning Toolbox*) for the SVR and MLP.

Table 6 Results of Machine Learning models on the test dataset

Set	Metric	Set Ph	Set A	Set L
SVRl	Corr (y_i, \tilde{y}_i)	0.82278	0.7601	0.836
	MAE*1e-3	0.48422	0.5486	0.47481
	MAPE (%)	19.24	23.24	19.10
SVRp	Corr (y_i, \tilde{y}_i)	0.68539	0.8274	0.828
	MAE*1e-3	0.57837	0.6314	0.62309
	MAPE (%)	25.20	28.26	27.08
MLP	Corr (y_i, \tilde{y}_i)	0.75059	0.90838	0.99356
	MAE*1e-3	0.6327	0.2006	0.05823
	MAPE (%)	30.52	8.89	2.67

For each algorithm, the prediction performance during the test period is shown in Table 6.

The results show clearly that the performance obtained by using the input set X_L (15) is better than that obtained with set X_A .

In particular, when the SVRl is applied, the set X_L allows for a 4.14 % reduction in MAPE, whereas, in the case of MLP, the MAPE improves by 6.22 %. Specifically, the predictive performance of MLP, using the set X_L , is highly competitive.

5 Conclusions

Among hydropower plants, those equipped with storage technology play a crucial role for the electric power system and for the energy transition. This technology allows water resources from watercourses or lakes to be stored in reservoirs, and this enables plant management to choose whether and when to release water to produce electricity. As a result, storage hydropower plants offer flexibility to the grid and help to mitigate the short-term production uncertainty that affects most green energy technologies. Hence, using water in reservoirs represents an opportunity cost, which is related to the evolution of production profitability and plant production capacity.

Due to these operational issues, predicting storage hydropower production requires addressing two problems of a different nature. On one side, a physical problem arises, i.e., predicting production capacity in the medium-term. On the other hand, an economic problem must be addressed, i.e., maximizing revenues by exploiting production capacity.

Regarding the economic issue, it is crucial to consider that, in a competitive power market, each producer's revenue depends on both the price of power and the generation supply of other price-maker producers.

Since the main price-makers in the power market are thermoelectric and hydroelectric producers, the economic variables to be used to predict hydropower generation are power prices and market values influencing thermoelectric production.

The main problem with incorporating these economic variables into a large-scale prediction model is that there are potentially too many types of variables to use. Thus,

it is important to consider that the strategies of market players are based on their short- and medium-term expectations. This implies that the problem can be simplified by using forward prices as a predictor.

In this paper we show that expectations on the Clean Spark Spread have an important impact on storage hydropower generation. In particular, for some time horizons, expectations on the *CSS* have a more important impact on hydropower generation than expectations on power price. Indeed, in these cases, the transfer entropy analysis shows a clear prevalence of the information flow from *CSS* to the *SHGN*, compared to the one in the opposite direction. This is because expectations of a lower *CSS* indicate that thermoelectricity will be offered at higher prices and vice versa.

Hence, there is an important effect on the overview of hydropower producers. In fact, the reduced competitiveness of thermal power plants will increase the share of demand that can be covered by storage hydropower generation. As a result, the future value of water in reservoirs increases and so does the current opportunity cost.

In addition, the insights provided by the transfer entropy analysis were used to identify the set of economic variables with the highest information content to predict *SHGN*. The results indicate that the subset of mixed prices X_L , identified according to the results of the TE analysis, is the best subset to predict *SHGN*. Specifically, the average conditional entropy of *SHGN* given X_L is 0.10873, which is a value significantly lower than the ones obtained using proxies of the expectations either on power price, $H(SHGN|X_A)$, or on *CSS*, $H(SHGN|X_B)$, that are respectively 0.135 and 0.1156.

Finally, we point out that it is of paramount importance to incorporate *CSS* expectations into the storage hydropower model. In fact, if the right mix of power price and *CSS* expectations is considered, the prediction error of the model is drastically reduced.

Specifically, in the case study we investigate, using an SVR with linear kernel we obtain a reduction in MAPE of 4.14% and 6,22% using MLP. The MLP algorithm based on set X_L , obtains a very competitive result in the problem, with a correlation of 0.99356 and a MAPE of 2.67%.

The methods employed here rely on two conditions, the stationarity (for TE method) and the iid assumption of the data (for the entropy in general). From Table 2, the ADF test rejects the null hypothesis of unit root also for all the series object of this article. This allows us to use the data in their raw form.

Anyway, local trends and seasonality could affect the robustness of the results. Nevertheless, the known approaches to address these issues, based on transformations of the series, would not be consistent with the economic theory underlying the model. For example, considering the series of differences to lag- k , for a suitable $k \geq 1$, may not be compatible with our paper contribution. In particular, the aim of our paper is to analyze the influence of expectations on hydroelectric generation forecasting, a topic that has not yet received enough consideration in the literature. By tackling this problem in terms of variations of storage hydropower generation, the economic link between dependent and independent variables would be lost. From an economic point of view, we would have no reason to argue that there is a relationship between

expectations on CSS in the medium term and the daily variation of hydroelectricity. This could be an interesting topic for future research.

Funding Open access funding provided by Università degli Studi Roma Tre within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A: Notation

See Table 7.

Appendix B: Transfer entropy based on quantile

We partitioned the data x_t into the discretized value s_t by the quantiles $\{q_i\}$ in Eq. (13). Specifically, we proceeded as in Park et al. (2021), by performing a uniform binning: for the 3-quantiles case, we have the first quantile below 0.33, the second quantile between 0.33 and 0.66 and the third quantile over 0.66; if it is expressed in four

Table 7 Notation

Notation	Description	Unit of measure
<i>SHGN</i>	Storage Hydropower generation in Northern Italy	Gwh
<i>D01</i>	Forward one Day Power Price	€/MWh
<i>W01</i>	Forward one Week Power Price	€/MWh
<i>MX_{pw}</i>	Forward X Mouthy Power Price	€/MWh
<i>QX_{pw}</i>	Forward X Quarter Power Price	€/MWh
<i>MX_{gas}</i>	Forward X Mouthy Gas Price	€/MWh
<i>QX_{gas}</i>	Forward X Quarter Gas Price	€/MWh
<i>MXC02</i>	Forward X Mouthy C02 Price	€/MWh
<i>QXC02</i>	Forward X Quarter C02 Price	€/MWh
<i>CSSD01</i>	CSS Forward one Day Power Price	€/MWh
<i>CSSMX</i>	CSS Forward X Mouthy Power Price	€/MWh
<i>CSSQX</i>	CSS Forward X Quarter Power Price	€/MWh
CE	Conditional Entropy	bits
ETE	Effective Transfer Entropy	bits

quantiles, the first quantile is below 0.25, the second quantile falls between 0.25 and 0.50, the third quantile is between 0.50 and 0.75, while the fourth quantile is over 0.75.

As noted by Park et al. (2021), when data is partitioned, the selection of a suitable number of quantiles is crucial. If this number is too big, the interval of interest will be too narrow. By contrast, if it is too small, the interval of interest will be too wide. For this reason, some established practices have been proposed in the literature (Benedetto et al. 2020; Behrendt and Schmidt 2020; Behrendt and Prange 2021). Moreover, since misdefining the number of quantiles can reduce TE and affect the validity of the analysis, Park et al. (2021) proposed to consider, as a suitable number of quantiles, those value that maximize the TE. Anyway, this approach based on uniform binning does not account for leptokurtic series and does not allow to determine the dominant direction of the information flow across different quantile choices. In fact, as explained by Behrendt et al. (2019): “the information flow in both directions cannot be compared across different quantiles but only for the same quantile”. However, we still performed the analysis of the transfer entropy between SHGN and the other series, by increasing the number of quantiles incrementally from 1 to 10, as done by Park et al. (2021).

Figures 9 and 10 show TE between SHGN and the rest of the dataset according to a function of number of quantiles q . When $q = 1$, TE is zero, because discretized data fall into only one process. As q increases, TE increases as well and, unlike Park et al. (2021) results, it is not maximized in both directions for the same value of q . Moreover, unlike Park et al. (2021) results, it is not true that, above $q = 3$, TE rapidly decreases and becomes close to zero. Nevertheless, this analysis confirms, for almost all the values of q , the results enclosed in the last column of 3. For example, the upper-left plot in Fig. 9 confirms that the dominant direction is from CSSD01 to SHGN for each value of q between 1 and 10. These figures shed light also on some doubtful findings returned by last column of Table 3. For example, Fig. 9 shows that, in the case of CSSM02 forward value, the dominant direction is from CSSM02 to SHGN.

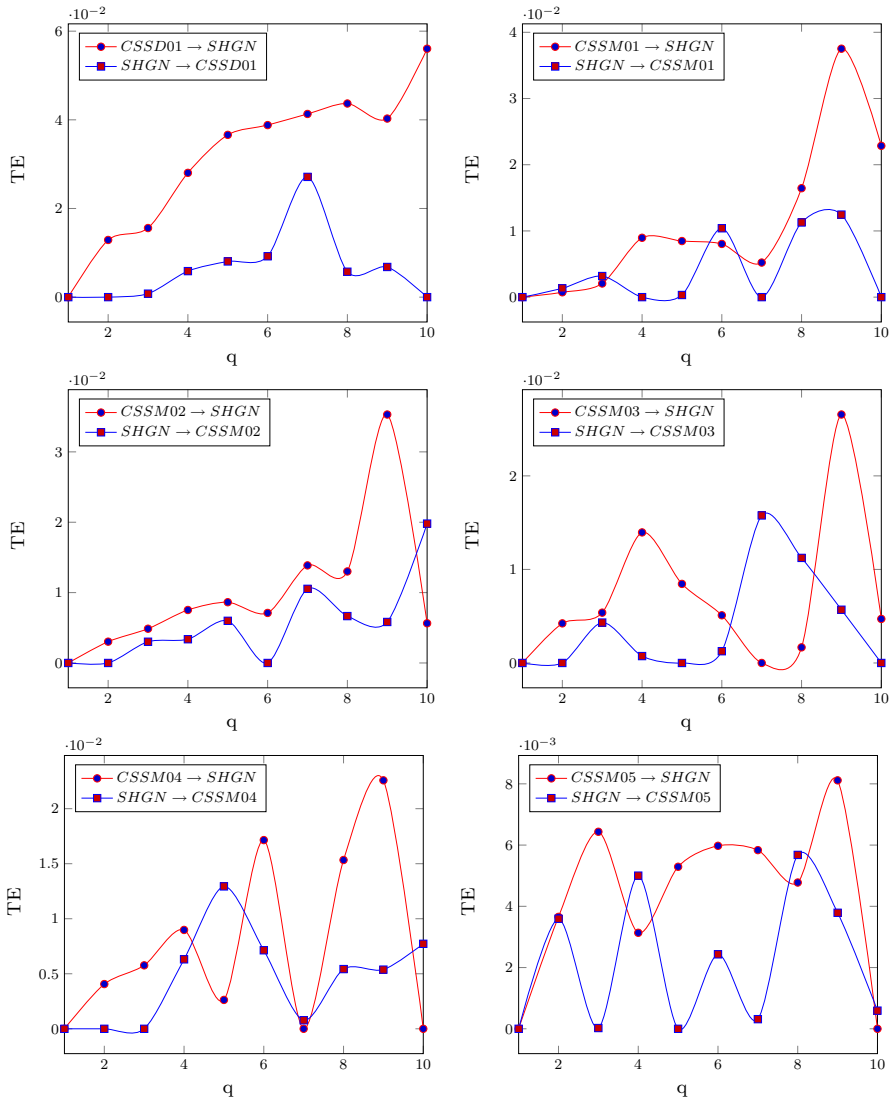


Fig. 9 Transfer entropy as a function of q : SHGN versus CSSD01, CSSM01, . . . , CSSM05

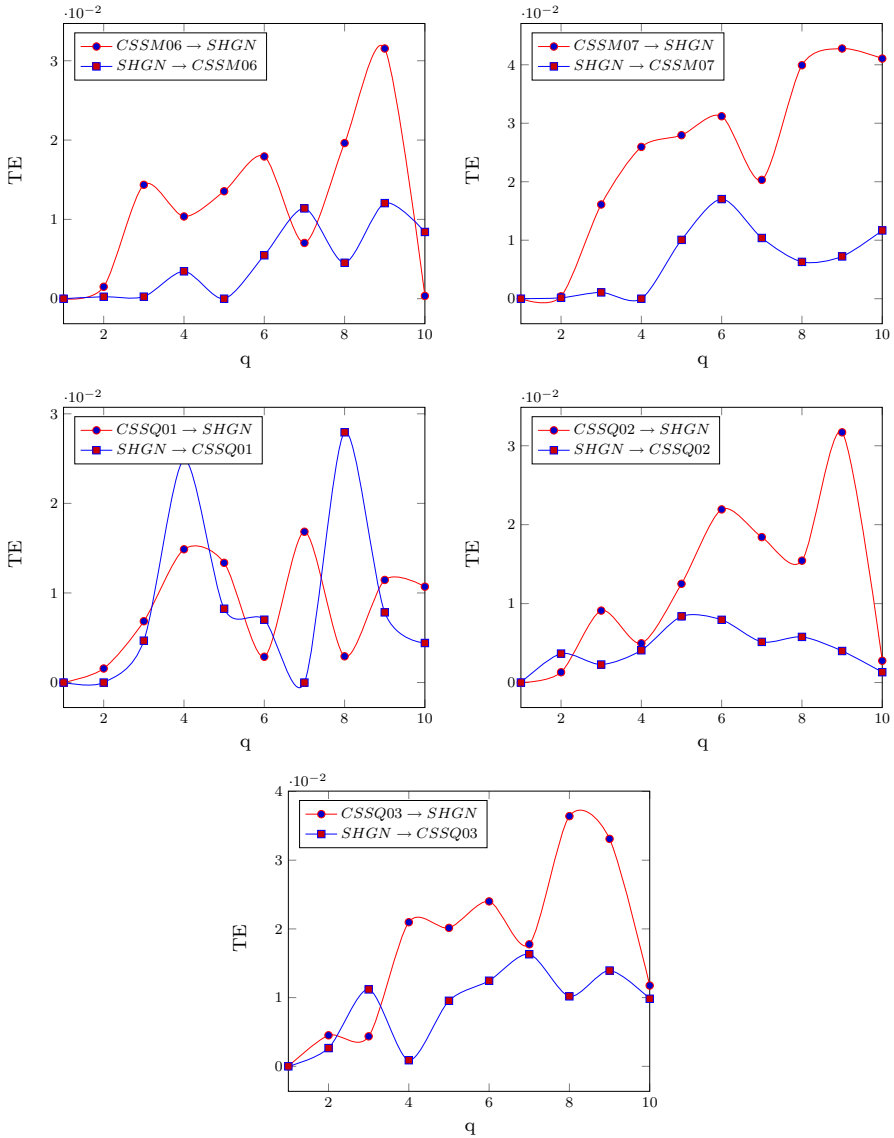


Fig. 10 Transfer entropy as a function of q : SHGN versus CSSM06, CSSM07, CSSQ01, CSSQ02, CSSQ03

Appendix C: plots of the most representative series

The series depicted in this Appendix corresponds to sub-hydrographic basins in Northern Italy depicted in Fig. 11. They are: Figs. 12, 13, 14, 15 and 16.

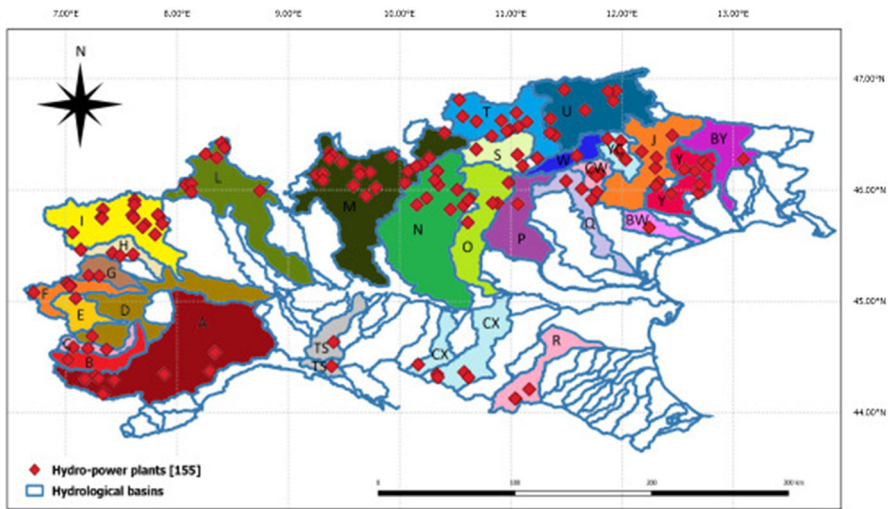


Fig. 11 Northern Italy sub-hydrographic basins and the grouping of hydro-power plants

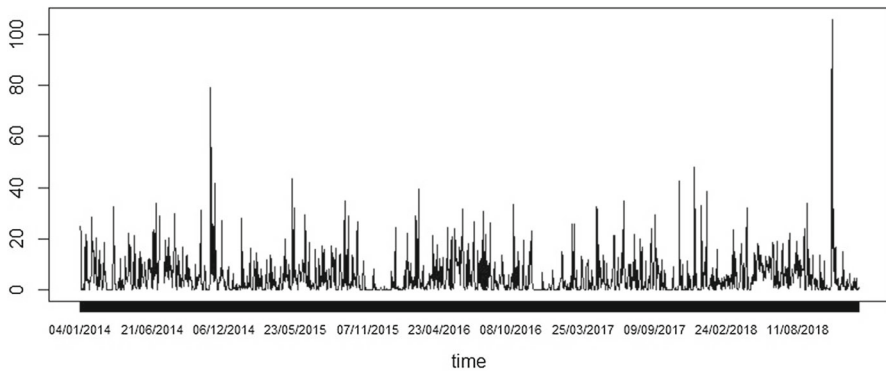


Fig. 12 Series of the rainfall on Alps-Italy (cumulative daily in millimeters)

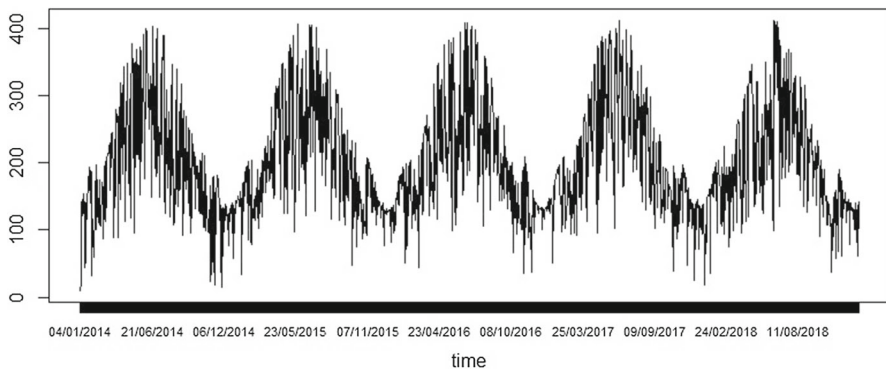


Fig. 13 Daily average global radiation MJ/m²

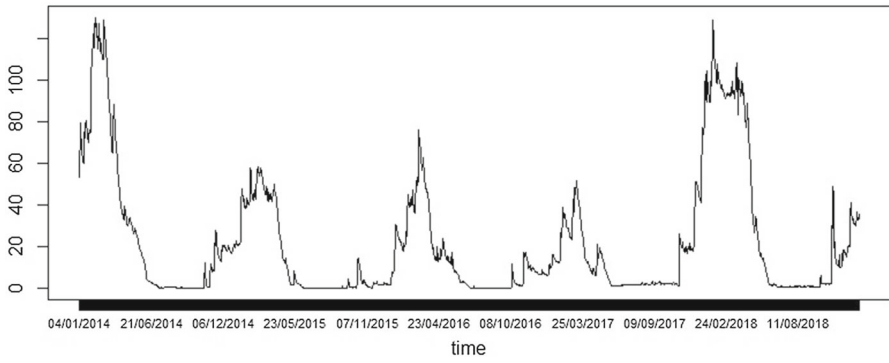


Fig. 14 Snowpack level (daily average in centimeters)

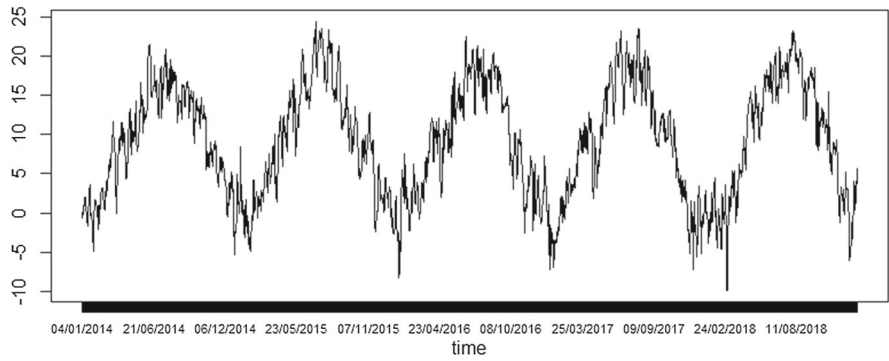


Fig. 15 Average daily temperature in °C

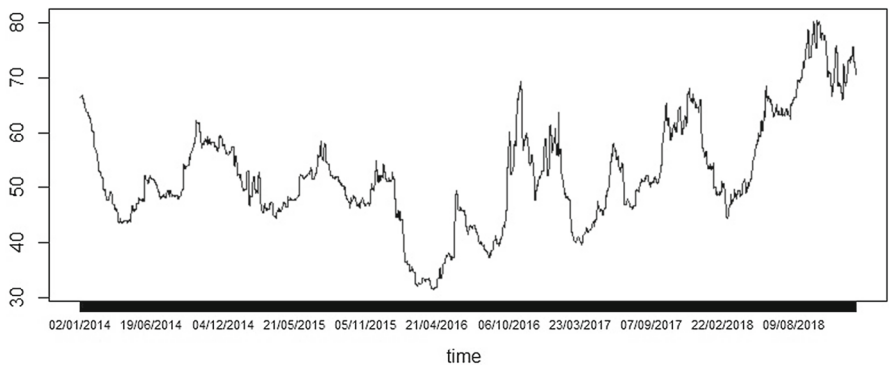


Fig. 16 The current price of the monthly forward with delivery period in the next calendar month, $M01$ (/MWh)

References

- Aasgård, E.K., Fleten, S.E., Kaut, M., Midthun, K., Perez-Valdes, G.A.: Hydropower bidding in a multi-market setting. *Energy Syst.* **10**(3), 543–565 (2019)
- Adnan, J., Daud, N.N., Mokhtar, A., Hashim, F., Ahmad, S., Rashidi, A., Rizman, Z.: Multilayer perceptron based activation function on heart abnormality activity. *J. Fundam. Appl. Sci.* **9**(3S), 417–432 (2017)
- Ahmad, S.K., Hossain, F.: A generic data-driven technique for forecasting of reservoir inflow: application for hydropower maximization. *Environ. Model. Softw.* **119**, 147–165 (2019)
- Albadi, M., El-Saadany, E.: Overview of wind power intermittency impacts on power systems. *Electr. Power Syst. Res.* **80**(6), 627–632 (2010)
- Assis, J., de Assis, F.: Estimation of transfer entropy between discrete and continuous random processes. *J. Commun. Inf. Syst.* **33**, 1–11 (2018)
- Banerjee, A., Dolado, J.J., Galbraith, J.W., Hendry, D., et al.: Co-Integration, Error Correction, and the Econometric Analysis of Non-stationary Data. OUP Catalogue, Oxford (1993)
- Baslis, C.G., Bakirtzis, A.G.: Mid-term stochastic scheduling of a price-maker hydro producer with pumped storage. *IEEE Trans. Power Syst.* **26**(4), 1856–1865 (2011)
- Behrendt, S., Prange, P.: What are you searching for? On the equivalence of proxies for online investor attention. *Finance Res. Lett.* **38**, 101401 (2021)
- Behrendt, S., Schmidt, A.: Nonlinearity matters: the stock price-trading volume relation revisited. *Econ. Model.* **98**, 371–385 (2020)
- Behrendt, S., Dimpfl, T., Peter, F.J., Zimmermann, D.J.: Rtransferentropy—quantifying information flow between different time series using effective transfer entropy. *SoftwareX* **10**, 100265 (2019)
- Benedetto, F., Mastroeni, L., Quaresima, G., Vellucci, P.: Does OVX affect WTI and Brent oil spot variance? Evidence from an entropy analysis. *Energy Econ.* **89**, 104815 (2020)
- Birkedal, M., Bolkesjø, T.F.: Determinants of regulated hydropower supply in Norway. *Energy Procedia* **87**, 11–18 (2016)
- Castillo-Botón, C., Casillas-Pérez, D., Casanova-Mateo, C., Moreno-Saavedra, L., Morales-Díaz, B., Sanz-Justo, J., Salcedo-Sanz, P., et al.: Analysis and prediction of dammed water level in a hydropower reservoir using machine learning and persistence-based techniques. *Water* **12**(6), 1528 (2020)
- Chen, D., Leon, A.S., Gibson, N.L., Hosseini, P.: Dimension reduction of decision variables for multireservoir operation: a spectral optimization model. *Water Resour. Res.* **52**(1), 36–51 (2016)
- Chen, N., Xiong, C., Du, W., Wang, C., Lin, X., Chen, Z.: An improved genetic algorithm coupling a back-propagation neural network model (IGA-BPNN) for water-level predictions. *Water* **11**(9), 1795 (2019)
- Condemì, C., Casillas-Pérez, D., Mastroeni, L., Jiménez-Fernández, S., Salcedo-Sanz, S.: Hydro-power production capacity prediction based on machine learning regression techniques. *Knowl. Based Syst.* **222**, 107012 (2021)
- Condemì, C., Mastroeni, L., Vellucci, P.: Selection of predictor variables to aggregate generation model. *J. Energy Mark.* **14**(1), 27–60 (2021)
- Cover, T.M., Thomas, J.A.: *Elements of Information Theory*. Wiley, Hoboken (2012)
- Cuo, L., Pagano, T.C., Wang, Q.: A review of quantitative precipitation forecasts and their use in short-to medium-range streamflow forecasting. *J. Hydrometeorol.* **12**(5), 713–728 (2011)
- Dimpfl, T., Peter, F.J.: Analyzing volatility transmission using group transfer entropy. *Energy Econ.* **75**, 368–376 (2018)
- Dougherty, J., Kohavi, R., Sahami, M.: Supervised and unsupervised discretization of continuous features. In: *Machine Learning Proceedings 1995*, pp. 194–202. Elsevier (1995)
- EEX: European energy exchange. <https://www.eex.com/en/market-data/power/futures> (2019). Accessed 3 Jan 2019
- Fischer, I., Alemi, A.A.: CEB improves model robustness. <https://openreview.net/forum?id=SygEukHYvB> (2020). Accessed 3 Jan 2020
- Friston, K., Adams, R., Perrinet, L., Breakspear, M.: Perceptions as hypotheses: saccades as experiments. *Front. Psychol.* **3**, 151 (2012)
- Ghoddusi, H., Creamer, G.G., Rafizadeh, N.: Machine learning in energy economics and finance: a review. *Energy Econ.* **81**, 709–727 (2019)
- He, J., Shang, P.: Comparison of transfer entropy methods for financial time series. *Physica A* **482**, 772–785 (2017)

- Hirth, L.: The benefits of flexibility: the value of wind energy with hydropower. *Appl. Energy* **181**, 210–223 (2016)
- IHA: Hydropower status report. International Hydropower Association: London, UK, Tech rep (2018)
- ISPRA: Fattori di emissione atmosferica di gas ad effetto serra e altri gas nel settore elettrico. Istituto Superiore per la Protezione e la Ricerca Ambientale, Tech rep (2018)
- ISTAT: Istituto nazionale di statistica. Territory and cartography database. <https://www.istat.it/en> (2020). Accessed 3 Jan 2021
- Jahns, C., Podewski, C., Weber, C.: Supply curves for hydro reservoirs-estimation and usage in large-scale electricity market models. *Energy Econ.* **87**, 104696 (2020)
- Kayri, M.: Predictive abilities of Bayesian regularization and Levenberg–Marquardt algorithms in artificial neural networks: a comparative empirical study on social data. *Math. Comput. Appl.* **21**(2), 20 (2016)
- Killingtveit, Å.: Chap 8: Hydropower. In: Letcher, T.M. (ed.) *Managing Global Warming*, pp. 265–315. Academic Press, Cambridge (2019)
- Kohavi, R., John, G.H.: Wrappers for feature subset selection. *Artif. Intell.* **97**(1–2), 273–324 (1997)
- Komiyama, R., Fujii, Y.: Assessment of massive integration of photovoltaic system considering rechargeable battery in Japan with high time-resolution optimal power generation mix model. *Energy Policy* **66**, 73–89 (2014)
- Kullback, S., Leibler, R.A.: On information and sufficiency. *Ann. Math. Stat.* **22**(1), 79–86 (1951)
- Lee, J., Nemati, S., Silva, I., Edwards, B.A., Butler, J.P., Malhotra, A.: Transfer entropy estimation and directional coupling change detection in biomedical time series. *Biomed. Eng. Online* **11**(1), 19 (2012)
- Li, G.D., Masuda, S., Nagai, M.: Prediction of hydroelectric power generation in Japan. *Energy Sour. Part B Econ. Plan. Policy* **11**(3), 288–294 (2016)
- Liu, H., Hussain, F., Tan, C.L., Dash, M.: Discretization: an enabling technique. *Data Min. Knowl. Discov.* **6**(4), 393–423 (2002)
- Meyer, P.E.: Information-theoretic variable selection and network inference from microarray data. PhD thesis, Université Libre de Bruxelles (2008)
- Meyer, P.E.: Package infotheo. Princeton, NJ, USA, R Package Version; Citeseer (2009)
- Mohd Yassin, I., Jailani, R., Megat Ali, M.S.A., Baharom, R., Abu Hassan, A.H., Rizman, Z.I.: Comparison between cascade forward and multi-layer perceptron neural networks for NARX functional electrical stimulation (FES)-based muscle model. *Int. J. Adv. Sci. Eng. Inf. Technol.* **7**(1), 215–221 (2017)
- Monteiro, C., Ramirez-Rosado, I.J., Fernandez-Jimenez, L.A.: Short-term forecasting model for aggregated regional hydropower generation. *Energy Convers. Manag.* **88**, 231–238 (2014)
- Moreno, J.: Hydraulic plant generation forecasting in Colombian power market using ANFIS. *Energy Econ.* **31**(3), 450–455 (2009)
- Mosavi, A., Salimi, M., Faizollahzadeh Ardabili, S., Rabczuk, T., Shamshirband, S., Varkonyi-Koczy, A.R.: State of the art of machine learning models in energy systems, a systematic review. *Energies* **12**(7), 1301 (2019)
- Muñoz, J.R., Sailor, D.J.: A modelling methodology for assessing the impact of climate variability and climatic change on hydroelectric generation. *Energy Convers. Manag.* **39**(14), 1459–1469 (1998)
- Nandalal, K., Bogardi, J.J.: *Dynamic Programming Based Operation of Reservoirs: Applicability and Limits*. Cambridge University Press, Cambridge (2007)
- Paninski, L.: Estimation of entropy and mutual information. *Neural Comput.* **15**(6), 1191–1253 (2003)
- Park, S., Jang, K., Yang, J.S.: Information flow between bitcoin and other financial assets. *Physica A* **566**, 125604 (2021)
- Plucinski, B., Sun, Y., Wang, S.Y.S., Gillies, R.R., Eklund, J., Wang, C.C.: Feasibility of multi-year forecast for the Colorado river water supply: time series modeling. *Water* **11**(12), 2433 (2019)
- Rastrow, A., Dredze, M., Khudanpur, S.: Adapting n-gram maximum entropy language models with conditional entropy regularization. In: 2011 IEEE Workshop on Automatic Speech Recognition Understanding, pp. 220–225 (2011)
- Reunanen, J.: Overfitting in making comparisons between variable selection methods. *J. Mach. Learn. Res.* **3**(Mar), 1371–1382 (2003)
- Schreiber, T.: Measuring information transfer. *Phys. Rev. Lett.* **85**, 461–464 (2000)
- Singh, V.K., Singal, S.: Operation of hydro power plants—a review. *Renew. Sustain. Energy Rev.* **69**, 610–619 (2017)
- SNPA: Sistema nazionale per la protezione dell'ambiente. Hydrographic database. <https://www.snambiente.it/chi-siamo/i-nodi-del-sistema/i-siti-web> (2019). Accessed 3 Jan 2019

- Steeger, G., Rebennack, S.: Strategic bidding for multiple price-maker hydroelectric producers. *IIE Trans.* **47**(9), 1013–1031 (2015)
- TERNA: Terna spa. <https://www.terna.it> (2019). Accessed 3 Jan 2019
- Uzlu, E., Akpınar, A., Öztürk, H.T., Nacar, S., Kankal, M.: Estimates of hydroelectric generation using neural networks with the artificial bee colony algorithm for Turkey. *Energy* **69**, 638–647 (2014)
- Wang, Z.X., Li, Q., Pei, L.L.: Grey forecasting method of quarterly hydropower production in china based on a data grouping approach. *Appl. Math. Model.* **51**, 302–316 (2017)
- Wen, L.Y., Min, F., Wang, S.Y.: A two-stage discretization algorithm based on information entropy. *Appl. Intell.* **47**, 1169–1185 (2017)
- Weron, R.: Electricity price forecasting: a review of the state-of-the-art with a look into the future. *Int. J. Forecast.* **30**(4), 1030–1081 (2014)
- Yang, Y., Webb, G.I.: Discretization for Naive–Bayes learning: managing discretization bias and variance. *Mach. Learn.* **74**(1), 39–74 (2009)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.