Applied Energy 293 (2021) 116891

Contents lists available at ScienceDirect

Applied Energy

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

An adaptive short-term forecasting method for the energy yield of flat-plate solar collector systems

Viktor Unterberger [a,b], Klaus Lichtenegger [a,c], Valentin Kaisermayer [a,b], Markus Gölles [a,b,*], Martin Horn [a,b]

[a] BEST - Bioenergy and Sustainable Technologies GmbH, Inffeldgasse 21/B, 8010 Graz, Austria
[b] Institute for Automation and Control, Graz University of Technology, Inffeldgasse 21/B, 8010 Graz, Austria
[c] FH JOANNEUM – University of Applied Sciences, Institute of Information Management, Eckertstraße 30i, 8020 Graz, Austria

ARTICLE INFO

Keywords:
Forecasting
Solar thermal
Adaptive
Flat-plate collectors
Experimental validation
Weather forecasts

ABSTRACT

The number of large-scale solar thermal installations has increased rapidly in Europe in recent years, with 70 % of these systems operating with flat-plate solar collectors. Since these systems cannot be easily switched on and off but directly depend on the solar radiation, they have to be combined with other technologies or integrated in large energy systems. In order to most efficiently integrate and operate solar systems, it is of great importance to consider their expected energy yield to better schedule heat production, storage and distribution. To do so the availability of accurate forecasting methods for the future solar energy yield are essential. Currently available forecasting methods do not meet three important practical requirements: simple implementation, automatic adaption to seasonal changes and wide applicability. For these reasons, a simple and adaptive forecasting method is presented in this paper, which allows to accurately forecast the solar heat production of flat-plate collector systems considering weather forecasts. The method is based on a modified collector efficiency model where the parameters are continuously redetermined to specifically consider the influence of the time of the day. In order to show the wide applicability the method is extensively tested with measurement data of various flat-plate collector systems covering different applications (below 200 °Celsius), sizes and orientations. The results show that the method can forecast the solar yield very accurately with a Mean Absolute Range Normalized Error (MARNE) of about 5 % using real weather forecasts as inputs and outperforms common forecasting methods by being nearly twice as accurate.

1. Introduction

In 2019, solar thermal heat represented one of the top three renewable sources driving climate protection, together with wind power and photovoltaics [1]. In this context, the number of large-scale solar heating installations has increased rapidly in European countries, e.g., in Denmark [2], but also worldwide in the last couple of years, leading to the installation of about 400 large-scale solar thermal systems (>350 kW_{th}, 500 m^2) by the end of 2019 [1]. In Europe, about 70 % of all solar thermal systems installed are flat-plate solar collectors [1], making them an important technology in this sector.

Even though the systems are mature and achieve high efficiency rates, they cannot be simply switched on and off but directly depend on the solar radiation. For this reason, these systems have to be combined with other technologies or integrated into large energy systems. In both cases, it has been shown that a significant potential exists for improvement when considering the future solar yield in their high-level controllers, e.g. energy management system (EMS), see [3], in order to optimally schedule heat production, storage and distribution for the near future, e.g. [4]. For example, by considering the predicted solar energy yield, the number of unnecessary operations performed by other heat production units (e.g. gas boilers) can be reduced, while managing the heat storage and the distribution in ways that always allow the full solar yield to be used, saving money and increasing the overall system performance. Furthermore, critical overheating problems in the solar plant can be eliminated since knowing the predicted solar yield enables night-cooling strategies to be applied well ahead of time (e.g., see [5]).

In view of practical applicability, the forecasting method of the future solar energy yield should meet three requirements:

* Correspondence to: Inffeldgasse 21/B, 8010 Graz, Austria.
E-mail addresses: viktor.unterberger@best-research.eu (V. Unterberger), klaus.lichtenegger@fh-joanneum.at (K. Lichtenegger), valentin.kaisermayer@best-research.eu (V. Kaisermayer), markus.goelles@best-research.eu (M. Gölles), martin.horn@tugraz.at (M. Horn).

https://doi.org/10.1016/j.apenergy.2021.116891
Received 16 December 2020; Received in revised form 1 March 2021; Accepted 28 March 2021
Available online 16 April 2021

1. *Simple implementation* – The methods should not require high computational effort or depend on third party software in order to be platform independent and easily implementable on commercially available controllers.
2. *Automatic adaption* – The methods should automatically adapt to variations over the year (e.g. seasonal changes) minimizing the re-parameterization effort.
3. *Wide applicability* – The methods should be able to be used to describe a large variety of different solar collector installations, regarding application (hot water and process heat), size, orientation and climate conditions.

Before investigating how these aspects are addressed by currently available forecasting methods for the solar energy yield the results of an extensive literature review is presented.

In general, forecasting the energy yield of solar thermal systems is in comparison to the electrical sector, see e.g. [6,7], a smaller and younger research field. In this field, two methods are commonly available for predicting the behaviour of solar thermal systems: one is built upon physically based models, the second is a rapidly growing field based on computational intelligence techniques (CIT), see [8].

Regarding the first method, several models of different complexity could generally be used to forecast the solar energy yield. In [9], a complex model based on coupled partial differential equations is presented for describing the behaviour of solar collector fields, achieving a high level of accuracy. Furthermore, in [10], an ordinary differential equation is used to simulate the thermal efficiency of a parabolic trough and flat-plate collector field. Finally, in [11], a simple algebraic collector model is used to perform an online forecast of the energy yield from solar thermal collectors. The advantages of these model-based methods are that high degrees of accuracy for small time steps, e.g. 1 s, can generally be obtained as well as their physically motivated modelling approach. The disadvantages are the necessity of having domain knowledge, the frequently high levels of complexity and the need to perform an exhaustive exploration of their (many) parameters.

Regarding the second method, several recent contributions use computational intelligence techniques (CIT) to predict the solar energy yield. In [12] artificial neural networks (ANNs) are used to predict the performance of large solar systems including the expected energy yield. In [13], an artificial neural network (ANN) was also used to predict the solar yield and local outside temperatures. A review on using ANN techniques to predict the performance of solar collector systems was presented [14]. Furthermore, in [15], ANN is compared to an adaptive neuro fuzzy interface system (ANFIS) to model and predict the efficiency of flat-plate solar collectors. ANFIS is also used in [16] for modelling the performance prediction of a whole solar thermal energy system. Besides neural-network-based algorithms, Ahmad et al. [8] uses machine learning algorithms to predict the useful solar thermal energy, by investigating support vector machines, random forest and extra trees. Finally, in [17], deep learning models are trained to predict the performance of solar hot water systems under different meteorological conditions, exploring ANN, recurrent neural networks and long short-term memory. The advantages of these computational intelligence techniques are that no domain knowledge is necessary, resulting in a black box modelling approach which can be easily applied to different systems. The disadvantages are the need for exhaustive data, the risk of over-fitting, the necessary computing power and the often platform dependent algorithms.

Furthermore, hybrid approaches exist that use a solar thermal model where the parameters are determined by applying CIT techniques. For example, this approach was performed to predict the collector performance in a similar way in [18–21].

Regarding the *simple implementation*, for the methods using physical models its depend on the model structure if they can be simply implemented. While the model from [9] needs experts for the implementation the models from [10,11] can be implemented simply. The methods based on machine learning approaches, including those described in [8,12,13,15–19], need significant computational power, often rely on third-party software, large amounts of data from several inputs and typically experts for their implementation as well as their parameterization. Thus these methods cannot be considered to be simply implemented in off-the-shelf controllers, which have strict limitations regarding their performance, memory and software.

Regarding the *automatic adaption*, only Bacher et al. [11] mention that the applied models adapt over time due to changes in the surroundings, wear or exposure to dirt, but these adaptations and changes were not verified in the provided simulation studies. The models described in [9,10] are not designed for an automatic adaption, as they have fixed parameters. Machine learning approaches, Kalogirou [18], Xie et al. [19], Khademi et al. [15], Kalogirou et al. [12], Yaïci and Entchev [16], Kramer and Bitterling [13], Ahmad et al. [8] and Correa-Jullian et al. [17], can be considered generally as adaptive; however, the adaptivity is often not specifically evaluated in the contributions and typically the additional training process requires a large amount of data and extensive computational resources.

Finally, regarding the *wide applicability*, the methods are developed and evaluated only for one single plant, Kalogirou [18], Xie et al. [19], Pasamontes et al. [9], Khademi et al. [15], Kalogirou et al. [12], Yaïci and Entchev [16], Kramer and Bitterling [13], Ahmad et al. [8] and Tian et al. [10], or merely in simulation studies, Bacher et al. [11], Lee et al. [21] and Correa-Jullian et al. [17]. Therefore, the wide applicability regarding application, size, orientation, or climate has not yet been shown.

Overall, this review of the literature reveals a research gap: No forecasting method currently exists that enables to predict the solar energy yield and meet the three above mentioned requirements of being simple to implement and therefore platform independent, automatically adapts itself to changing conditions and is widely applicable. Only Nigitz and Gölles [22] proposes a simple, adaptive and widely applicable method that can be used to forecast the consumers' heat demand. Since, this method meets the aforementioned three requirements, there has been already a first incomplete attempt of the authors to generalize this method in order to use it also for the heat production of collector installations [23]. However, this first incomplete attempt of the authors was the basis to fully develop, improve and comprehensively evaluate the method leading to this paper. Among other aspects is the method improved regarding: structure of the correction function, determination of its model parameters, evaluation of different optimization algorithms, extensive evaluation of its short-term and long-term accuracy with real measurements of a multiplicity of solar systems, practical applicability regarding its long-term accuracy using real weather forecasts, complete description of the method and its advantages over other forecasting methods, extensive investigation of all its parameters and recommendations on how to choose them, comparison to two currently available commonly used forecasting methods.

The research gap is filled by the novel method since it is based on a modified efficiency equation of flat-plate solar collectors where all parameters are continuously redetermined to specifically consider the influence of the time of the day and seasonal changes. The adaption is based on linear regression approach and therefore allows a platform independent implementation. Finally, since the method is based on a modified energy balance that is valid for all kind of flat-plate collectors, it can be assumed to be widely applicable. This wide application is proved by performing an comprehensive practical validation with real measurement data from several collector systems.

The paper is structured as follows: in Section 2, the forecasting method is derived from two forecasting principles that are often used in the literature: the seasonal naïve method presented in Section 2.1 and the data sheet method presented in Section 2.2. The resulting forecasting method, presented in Section 2.3, is then experimentally validated in Section 3, including a description of the experimental setup in Section 3.1, a discussion of validation measures in Section 3.2, the parameters determination in Section 3.3 and the experimental results in Section 3.4. Finally, in Section 4, a conclusion about the presented method is drawn and a short outlook is provided.

2. Forecasting method

First, two already existing and principally suitable methods for forecasting the solar yield are investigated and their advantages and disadvantages are discussed: the seasonal naïve method is presented in Section 2.1 and the data sheet method is presented in Section 2.2. Based on the findings of these discussions, the new adaptive forecasting method is derived and presented in Section 2.3; this method combines the advantages of the two aforementioned methods and, at the same time, eliminates their disadvantages. The superiority of the newly developed adaptive method is later shown in Section 3, whereby the two available methods are used as benchmarks for the performance evaluation.

2.1. Intuitive forecast — the seasonal naïve method

The sun follows approximately the same apparent path in the sky every day as the day before and approximately the same amount of heat will be generated at the same time every day, provided that the weather has not significantly changed. Thus, this natural 24-h-periodicity is a reasonable starting point for developing any forecasting method to determine solar yield. In a first approximation, one can predict the solar heat production at a particular time of the day as being the same as the production measured on the previous day at the same time

$$\hat{\dot{Q}}_{\mathrm{coll}}(t) = \dot{Q}_{\mathrm{coll}}(t - 24\,\mathrm{h}), \tag{1}$$

with $\hat{\dot{Q}}_{\mathrm{coll}}$ being the forecast for the solar heat for a time t in the future and $\dot{Q}_{\mathrm{coll}}(t - 24\,\mathrm{h})$ being the measured value of the solar heat at the same time on the previous day. This approach is known in literature as the *persistence model*, e.g. [6], or the *naïve seasonal prediction method*, e.g. [24], as it will be also called in this paper. This intuitive forecast made by applying the seasonal naïve prediction method works well if approximately the same conditions (and, in particular, similar amounts of solar radiation and similar ambient temperature) are measured on the subsequent day as on the previous one.

The advantages of this method are that it provides good results if the climate conditions remain constant, requires no parameterization, needs no external weather informations and adapts to seasonal changes. The method's biggest disadvantage is that the results can be highly incorrect if the conditions on the previous day are significantly different from those on the next one, which can lead to large deviations, especially on a short-term basis.

2.2. Static energy balance forecast — the data sheet method

While the seasonal naïve method can be considered as a black box method, it can make sense to take the physical behaviour of flat-plate collector systems into account for forecasting the solar yield. For this reason, the operating principle for solar heat production in these systems should be briefly reviewed before this principle is applied for forecasting purposes: The energy flows for a flat-plate collector are the heat input from the sun $\dot{Q}_{\mathrm{in}}$, the optical losses through reflection $\dot{Q}_{\mathrm{re}}$, the ambient heat losses through heat conduction and convection $\dot{Q}_{\mathrm{l,cc}}$ and the heat losses through radiation $\dot{Q}_{\mathrm{l,r}}$. Consideration of these flows leads to the following energy balance equation:

$$\dot{Q}_{\mathrm{coll}}(t) = \dot{Q}_{\mathrm{in}}(t) - \dot{Q}_{\mathrm{re}}(t) - \dot{Q}_{\mathrm{l,cc}}(t) - \dot{Q}_{\mathrm{l,r}}(t). \tag{2}$$

These energy flows are shown in Fig. 1. The heat produced by a flat-plate collector $\dot{Q}_{\mathrm{coll}}$ during steady-state operating conditions can be approximately expressed by the static energy equation according to the European Standard EN 12975-2; section 6.1, e.g. [25]:

$$\begin{aligned}\dot{Q}_{\mathrm{coll}}(t) = {} & A_{\mathrm{coll}} K(\theta) \eta_0 I_{\mathrm{g}}(t) - A_{\mathrm{coll}} c_1 \left(\bar{T}_{\mathrm{fl}}(t) - T_{\mathrm{amb}}(t)\right) \\ & - A_{\mathrm{coll}} c_2 \left(\bar{T}_{\mathrm{fl}}(t) - T_{\mathrm{amb}}(t)\right)^2,\end{aligned} \tag{3}$$

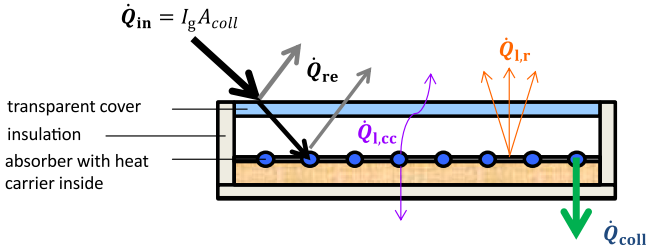

Fig. 1. Schematic structure of a flat-plate collector and the energy flows $\dot{Q}$ that occur: the heat input from the sun $\dot{Q}_{\mathrm{in}}$, the optical losses through reflection $\dot{Q}_{\mathrm{re}}$, the ambient heat losses through heat conduction and convection $\dot{Q}_{\mathrm{l,cc}}$ the heat losses through radiation $\dot{Q}_{\mathrm{l,r}}$ and the heat produced by the flat-plate collector $\dot{Q}_{\mathrm{coll}}$.

where A_{coll} denotes the net collector area, I_{g} the global radiation received by the collector surface, $\bar{T}_{\mathrm{fl}}$ the arithmetic mean fluid temperature between the inlet and the outlet of the collector and T_{amb} the ambient temperature. The coefficients represent the optical efficiency η_0, and, the heat loss coefficients for heat conductance, c_1, as well as for thermal radiation, c_2. The function $K(\theta)$ represents the incident angle modifier (IAM) which describes the dependency of the optical efficiency η_0 on the angle of incidence θ of the global solar radiation I_{g}. This angle varies from collector to collector and is typically estimated by performing experiments [26] and given in the data sheet.

In order to use Eq. (3) to forecast the solar yield $\hat{\dot{Q}}_{\mathrm{coll}}(t)$ for a time t in the future, forecasts for the following variables are necessary: $\hat{\bar{T}}_{\mathrm{fl}}$, $\hat{I}_{\mathrm{g}}$ and $\hat{T}_{\mathrm{amb}}$, which have to be estimated ($\hat{\bar{T}}_{\mathrm{fl}}$) or obtained from a weather service provider ($\hat{I}_{\mathrm{g}}$, $\hat{T}_{\mathrm{amb}}$).

In conclusion, the advantages of using the static energy balance of the collector as forecasting method for the solar yield are that this method considers the physical behaviour of the system, can be parameterized by the data sheet and can take into account the forecasts for external factors (e.g. $\hat{T}_{\mathrm{amb}}$) provided by a weather service provider. The disadvantages of the method are that even though the Standard EN12975:2006 is accepted and widely used, the results of analysing measurement data from solar thermal plants shows that applying this model, with its parameters (A_{coll}, η_0, c_1, c_2) taken from the data sheet of the collector, does not always lead to satisfying results for forecasting the solar heat production even if the external factors I_g, T_{amb} and operating conditions $\bar{T}_{\mathrm{fl}}$ are known by measurements. This is also confirmed by the experimental validation presented in Section 3. This has two important reasons:

1. The influence of the external factors ($I_{\mathrm{g}}, T_{\mathrm{amb}}$) determined by the coefficients (A_{coll}, η_0, c_1, c_2) would have to change over the day in order to consider the conditions occurring during daily operation, e.g. the thermal inertia, since some solar energy is needed to heat up the thermal masses of the collector in the morning, resulting in a lower solar heat production. Furthermore, systematic shading might be also relevant at some times of the day.
2. The model parameters ($\eta_0, c_1, c_2, A_{\mathrm{coll}}$) would not only have to change over the day but also over the entire time of operation to account for real world circumstances, such as dirty collector surfaces, which decrease the optical efficiency (given by η_0); material ageing, which leads to higher heat losses of the collectors (given by c_1 and c_2); and seasonally dependent local shading, which reduces the effective collector area (given by A_{coll}).

This means that neither the naïve nor the data sheet method yields fully satisfying results, and each of them has its shortcomings. For this reason, the new adaptive method presented here was developed to combine the advantages of both methods (naïve: adaptive, no parameterization effort, consideration of the past; data sheet: physically motivated, consideration of future weather information) and to eliminate their disadvantages.

2.3. New adaptive forecasting method

In order to combine the advantages of both methods and eliminate their disadvantages the new forecasting should

- be based on a physical model in order to be easily understandable,
- consider external factors (e.g. I_g, T_{amb}) as these can be nowadays obtained easily and accurately from weather service providers,
- behave potentially differently over the day in order to take into account short-term effects, such as thermal inertia or local shading,
- adapt its behaviour automatically over the years of operation by learning from the past without the need for any manual interventions and
- react quickly if a deviation occurs between the forecast and the measured variable.

To meet all these requirements, the static energy equation (3) is adapted in Section 2.3.1 such that it is more suitable for forecasting the solar yield by using external factors, while still taking into account the time dependent behaviour. Then, in Section 2.3.2, a parameterization routine is described that allows the method to automatically adapt its parameters based on past measurements. Finally, in Section 2.3.3 a correction approach is introduced that allows the method to quickly react by taking the latest error into account.

2.3.1. Prediction by considering external factors

As stated earlier, accurate forecasts from external factors ($\hat{I}_g$ and $\hat{T}_{amb}$) can easily be obtained from weather service providers. This is not the case for forecasts of the mean fluid temperature $\hat{\bar{T}}_{fl}$, which depends on the application and its operation. However, since solar thermal collectors are typically operated with a constant desired outlet temperature that is more or less ensured by the use of a temperature controller, the outlet temperature of the solar collector can be assumed to be relatively constant. Furthermore, since the solar collectors of large-scale solar thermal plants are typically connected to the lowest part of a buffer storage, the inlet temperature can be assumed to be constant as long as the buffer is of a reasonable size and not inefficiently operated. And even though the inlet and desired outlet temperature may change over the year, one can usually assume that it will remain constant for over the next 24 h, which corresponds to a typical forecast horizon. By making these two simplifications, a constant in- and outlet temperature can be assumed for forecasting, leading to a constant mean fluid temperature $\hat{\bar{T}}_{fl}(t) = \hat{\bar{T}}_{fl} = \text{const}$. Introducing the variable temperature difference $\Delta\hat{T}(t)$ as

$$\Delta\hat{T}(t) = \hat{\bar{T}}_{fl} - \hat{T}_{amb}(t)\,, \tag{4}$$

Eq. (3) becomes

$$\hat{\dot{Q}}_{coll}(t) = A_{coll} K(\theta)\eta_0 \hat{I}_g(t) - A_{coll} c_1 \Delta\hat{T}(t) - A_{coll} c_2 \Delta\hat{T}(t)^2. \tag{5}$$

Furthermore, since the application in EMS only requires predictions to be made at discrete time steps, it is reasonable to discretize (5) by introducing a sampling time Δt. The sampling time is chosen as $\Delta t = 1$ h, since weather service providers typically provide their predictions for external factors, global radiation and ambient temperature for every full hour, e.g. [27]. Therefore, this yields to

$$\hat{\dot{Q}}_{coll}[k] = A_{coll} K(\theta)\eta_0 \hat{I}_g[k] - A_{coll} c_1 \Delta\hat{T}[k] - A_{coll} c_2 \Delta\hat{T}[k]^2 \tag{6}$$

for the discrete time values $k = n+1, \ldots, n+N_f$ with the discrete time variable $n = t/\Delta t$ describing the hour at which the prediction is carried out and with N_f denoting the number of future hours for which the solar yield $\hat{\dot{Q}}_{coll}$ should be forecast.

As mentioned in Section 2.2, the coefficients $(A_{coll}, \eta_0, c_1, c_2)$ which weight the influence of the external factors, can change over the day for many reasons (e.g., thermal inertia, local shading). In order to model this time-dependent behaviour, different coefficients are used for different (discrete) times of the day. Furthermore to reduce the number of parameters to a minimum, the coefficients are combined to form the following *model parameters*: $\beta_1 = A_{coll}\, K(\theta)\, \eta_0$, $\beta_2 = A_{coll}\, c1$ and $\beta_3 = A_{coll}\, c_2$, leading to

$$\hat{\dot{Q}}_{coll}[k] = \beta_1[k]\hat{I}_g[k] - \beta_2[k]\Delta\hat{T}[k] - \beta_3[k]\Delta\hat{T}[k]^2, \tag{7}$$

where the dependency of β_1 on $K(\theta)$ is implicitly considered since β_1 changes over the day as the position of the sun changes, causing different angles of incidence $K(\theta)$. This approach also allows to consider the switch-off times caused by the underlying low-level control system by simply setting the parameter sets for a time j to zero: $\beta_1[j] = \beta_2[j] = \beta_3[j] = 0$.

Considering the natural 24-h-periodicity as for the naïve method described in Section 2.1, it can be assumed that the model parameter β_j weighting the influence of the external factors at each time of the day will only slightly change from day to day, while the external factors can vary greatly. Therefore, Eq. (7) can be rewritten as

$$\hat{\dot{Q}}_{coll}[k] = \beta_1[m]\hat{I}_g[k] - \beta_2[m]\Delta\hat{T}[k] - \beta_3[m]\Delta\hat{T}[k]^2\,, \tag{8}$$

with the individual times of the day $m = k \bmod 24$ (for a sampling time of $\Delta t = 1$ h). After dividing the model parameters β_j into different parameter sets $\beta_j[m]$, which are valid for a certain hour of the day m, these have to be determined from historical data. However, in order to always use the latest information and adapt to seasonal changes the model parameters should be continuously re-determined on-line at every hour n. Thus the sets of model parameters should change over time, leading to:

$$\hat{\dot{Q}}_{coll}[k] = \beta_1[m|n]\hat{I}_g[k] - \beta_2[m|n]\Delta\hat{T}[k] - \beta_3[m|n]\Delta\hat{T}[k]^2. \tag{9}$$

Finally, since the forecasts of the external factors $\hat{I}_g, \hat{T}_{amb}$ (and therefore $\Delta\hat{T}$) will also typically be continuously updated at every hour n by the weather service provider in a practical implementation, it is reasonable to always use the most recent information. This leads to the prediction $\hat{\dot{Q}}_{coll}$ for the next N_f time steps at the hour n by

$$\hat{\dot{Q}}_{coll}[k|n] = \beta_1[m|n]\hat{I}_g[k|n] - \beta_2[m|n]\Delta\hat{T}[k|n] - \beta_3[m|n]\Delta\hat{T}[k|n]^2. \tag{10}$$

In summary, the application of the new adaptive forecasting method results in 24 so called *multiple linear regression* models and considers forecasts of the global radiation and the ambient temperature as external factors. In principle, it could be possible to consider even more external factors, such as wind speed or wind direction. However, these variables have comparatively little influence and are rarely measured at real plants. Furthermore, using more inputs would also increase the complexity, require more memory storage and make the application of the method more expensive, since each individual weather forecast quantity must be typically paid for. In addition, in case the collectors would be mounted on a tracking system, the method could be used as well, but both the measured and predicted solar radiation $\hat{I}_g$, would have to be correctly converted to the tracking surface.

The method is equivalent to the static collector model and, therefore, is valid for use in a wide range of collectors. Instead of having to rely solely on the parameters given in the data sheet, however, the collector parameters should be continuously adapted by measurement data in this approach, as is described in the next Section 2.3.2. The dependency on the time of day is taken into account by using different parameter sets for each time m of the day and, in this case every hour. This way, influencing factors as pollution of the collector fields and material decay are automatically incorporated. The effects of local shading, which would have to be incorporated into the static collector model by performing complicated 3D modelling

and shadow calculations, e.g. [26], are also automatically included and no manual parameterization is necessary. In addition, later enhancements, such as added solar collectors or changing environmental conditions (e.g. newly built buildings that shade the collectors), are also automatically incorporated into the forecasting process.

2.3.2. Continuous reparametrization

The continuous reparametrization is done by using historic measurement data and considering the natural 24-h-periodicity. For this reason, at every full hour n the model parameters β for that specific time of the day, $m = n \bmod 24$, are re-determined, using historical data. These historical data, referred to as training data in the following, consist of measurement data for of the independent variables X, with the external variables I_g as well as ΔT, which are calculated in this case with the measured mean fluid temperature $\bar{T}_{f1}$, and the dependent variable Y, with the measured solar yield $\dot{Q}_{coll}$ for a certain number of training days N_d in the past. For example, in order to re-determine the set of model parameters at time n, meaning $m = n$, where n is aligned with the time of the clock, meaning that $n = 0$ corresponds to 00:00:00 of the first day considered, the following set of training data is used:

$$\boldsymbol{X}[n] = \begin{bmatrix} I_g[n] & I_g[n-1] & \dots & I_g[n-N_d] \\ \Delta T[n] & \Delta T[n-1] & \dots & \Delta T[n-N_d] \\ \Delta T[n]^2 & \Delta T[n-1]^2 & \dots & \Delta T[n-N_d]^2 \end{bmatrix} \tag{11}$$

$$\boldsymbol{Y}[n] = \begin{bmatrix} \dot{Q}_{coll}[n] & \dot{Q}_{coll}[n-1] & \dots & \dot{Q}_{coll}[n-N_d] \end{bmatrix}^\top . \tag{12}$$

How to choose the optimal number of training days N_d will be discussed in Section 3.3. However, in order to keep the method robust by taking days with different characteristics for training, the typical number of training days is larger than the number of model parameters β_j, leading to an overdetermined system of linear equations for the model parameters

$$\boldsymbol{Y}[n] = \underbrace{\boldsymbol{X}[n] \overbrace{\begin{bmatrix} \beta_1[n] \\ \beta_2[n] \\ \beta_3[n] \end{bmatrix}}^{\boldsymbol{\beta}[n]}}_{\hat{\boldsymbol{Y}}[n]} + \boldsymbol{e}[n] , \tag{13}$$

with $\boldsymbol{e}$ as the residual vector describing the error between the true value $\boldsymbol{Y}[n]$ and its estimation $\hat{\boldsymbol{Y}}[n]$. This means that the set of model parameters $\boldsymbol{\beta}$ depends on the way the residual vector $\boldsymbol{e}$ is minimized, which is indirectly defined by choosing a specific cost function J. For simplicity, J is chosen here as the sum of squared errors

$$J = \sum_i e_i^2 , \tag{14}$$

which leads to an ordinary least squares problem (OLS) for the optimal model parameters $\boldsymbol{\beta}$. The main advantage of this choice for J is that the OLS can be easily implemented by using the pseudo-inverse an, therefore, it can also be deployed on standard programmable logic controllers (PLCs). One disadvantage is that constraints cannot be considered in the optimization step. This means that the determination of the parameters can also yield to physically unreasonable values for the model parameters β, e.g. 'positive' heat losses. In such a case, the specific value for β is set to zero, and the calculation is performed again for the remaining model parameters. It has been shown that using a least squares optimization algorithm that can consider constraints, typically leads to better forecasting results, as exemplarity shown in Section 3.4.1. However, since not every target platform offers the possibility to use such a solver, the pseudo-inverse was used, since this is more generally applicable. Up until now, this forecasting method only uses measurement data from the last days, but does not take into account more recent measurements, thus incorrectly assuming that one time step is independent from the next. This would mean that, for example, the prediction made for 12:00 of the following day could be calculated on the current day at 12:00 (assuming a time step of one hour) and would not change afterwards. In order to reduce this shortcoming, a correction step is introduced which will be described in the next section.

2.3.3. Correction based on latest prediction error

A correction should help to increase the accuracy of the subsequent forecast of the near future by making use of the latest prediction error. Such a prediction error can occur for many reasons, one important reason being local temporary shading by clouds. Systematic effects may also occur: If it has rained over night, this may have increased the optical efficiency of the collector as compared to the efficiency on the previous days.

Such a prediction error is likely to persist over a certain time and detailed investigations of the prediction model with measurement data from solar heat producers have shown that the sign of the prediction error $(\dot{Q}_{coll} - \hat{\dot{Q}}_{coll})$ tends to stay the same over several time steps. Thus the last prediction error $(\dot{Q}_{coll}[n] - \hat{\dot{Q}}_{coll}[n|n-1])$ from the current time step n can be used to improve the prediction for the following time step $\hat{\dot{Q}}_{coll}[k|n]$, with $k > n$. Since the sign of the prediction error typically stays the same only over a few time steps, the correction is limited to the near future. In order to keep the method simple, a scaled linearly decaying function $\Phi[k]$ is used which adds the weighted current prediction error $(\dot{Q}_{coll}[n] - \hat{\dot{Q}}_{coll}[n|n-1])$ to the following few time steps, correcting the prediction. The correction can therefore be written as

$$\tilde{\dot{Q}}_{coll}[k|n] = \hat{\dot{Q}}_{coll}[k|n] + (\dot{Q}_{coll}[n] - \hat{\dot{Q}}_{coll}[n|n-1])\Phi[k|n] , \tag{15}$$

with Φ as linearly decaying correction function

$$\Phi[k|n] = \begin{cases} c_c \left(1 - \dfrac{k-n-1}{N_c - 1}\right) & \text{if } n+1 \le k \le n+N_c, \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

With the correction gain $c_c \in [0, 1]$ and the discrete time correction horizon given by $N_c \in \{2, \dots, N_f\}$.

While the correction (15) makes sense in general, it can also be the source of problems. For example can the solar yield exhibit an oscillating behaviour on an unsettled day (which of course also depends on the performance of the low-level control strategy of the plant). In such cases the assumption of the prediction error having the same sign for several time steps is violated, and consequently, the correction can actually worsen the prediction. This should be taken into account and evaluated when the method is applied in a specific application. However, for the measurement data described in 3.1, the sign of the prediction error is more likely to persist over a certain time. Thus, the correction on average improves the prediction, as will be shown in Section 3. The whole new adaptive forecasting method is illustrated in the flowchart shown in Fig. 2 for a sampling time of one hour.

3. Experimental validation

The adaptive forecasting method introduced in Section 2.3 should now be experimentally validated with real measurement data for several settings of flat-plate collector installations in order to evaluate its wide applicability. These settings together with the information regarding the installed measurement equipment are described in the first part of Section 3.1. Then in Section 3.2 measures used for the evaluation are chosen, and it is discussed why these are suitable for the purpose of evaluating the performance of the method. Furthermore, in Section 3.3, the few parameters needed to apply the adaptive forecasting method are determined. Finally, the validation results for different collector settings are shown in Section 3.4, considering different time ranges, the comparison to other forecasting methods and the use of real weather forecast data as inputs.

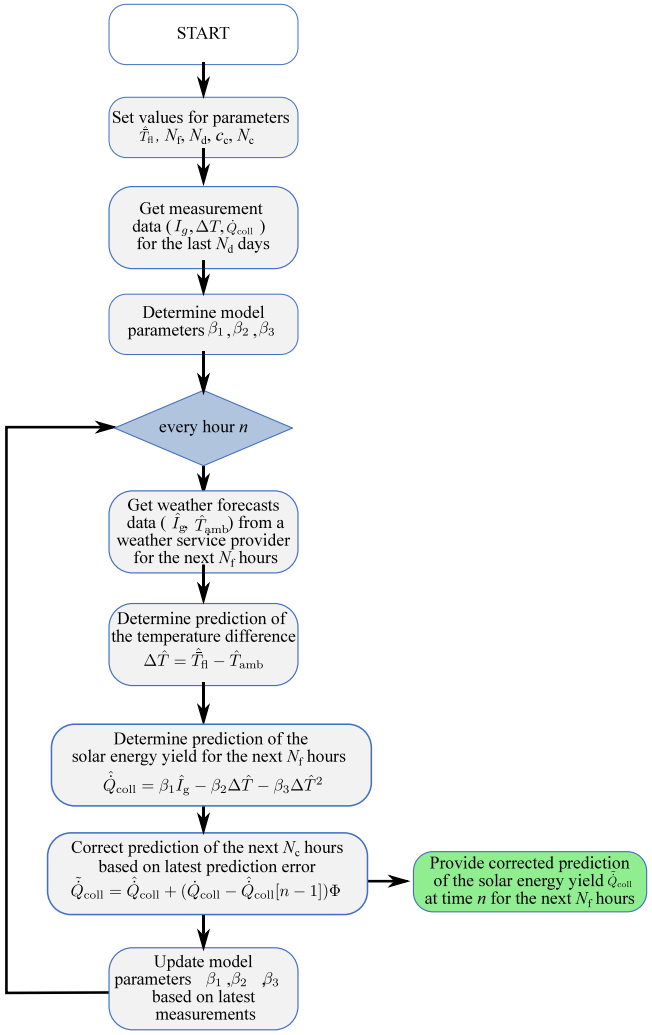

Fig. 2. Flowchart of the new adaptive forecasting method for predicting the solar energy yield.

3.1. Plant settings and measurements

3.1.1. Plant settings

The performance of the method is investigated for different flat-plate collector plants regarding several important aspects: application (hot water and process heat), size (from 200 m^2 to 3500 m^2), orientation (south, south-west, south east, 'mixed') and climate conditions (Austria and Kuwait) in order to demonstrate the wide applicability of the method. Regarding the climate conditions two strongly different but representative countries have been chosen: Austria with a medium to high solar radiation level and Kuwait with a high solar radiation. These collector systems with their different characteristics are summarized in Table 1, ordered by their nominal solar heat production. All fields in Austria are tilted by 45°, and the field in Kuwait by 15°; furthermore, their orientation from the south is given as value in the column *orientation* described by the surface azimuth angle γ, see e.g. [26], with zero for south, east negative, and west positive. The hot water applications have a desired outlet temperature of 80 °C and the process heat application of 160 °C.

For each field the solar yield is calculated by

$$\dot{Q}_{SH} = \dot{m}\, c_p(x, \bar{T})\, (T_{in} - T_{out}), \tag{17}$$

with the mass flow $\dot{m}$, the specific heat capacity of the fluid running through the collector c_p, which depends on the mass fraction of antifreeze medium x (for pure water $x = 0$) and the mean temperature $\bar{T}$ between the temperature entering T_{in} and leaving the collector field T_{out}. The mass flow $\dot{m}$ itself is calculated by

$$\dot{m} = \dot{V}\rho(x, T), \tag{18}$$

with the measured volume flow $\dot{V}$ and the density of the heat carrier ρ, depending on the mass fraction of glycol x and the temperature measured near the volume flow sensor T. The plant in Kuwait is operated with water as a heat carrier, while the plants in Austria are operated with an antifreeze mixture of $x = 35\,\%$ of the antifreeze medium *Tyfocor L* [28]. The values of heat capacity and density for the antifreeze mixture were taken from the data sheet and linearly interpolated for different temperature levels.

3.1.2. Measurements

The global solar radiation I_g on the tilted flat-plate collector surfaces was measured at all plants by using a pyranometer. In the plants in Austria, a Kipp&Zonen SMP6 [29] was used, and a MTX PCTRA056 [30] was used in Kuwait. Both are first class pyranometers, according to the classification in ISO 9060:1990 with an uncertainty of $< \pm 1\,\%$ and $< \pm 1.5\,\%$ respectively, regarding their long-term instability. The flow was measured in the Austrian plants with magnetic inductive flow sensors Optiflux400 from Krohne [31], which have an uncertainty of $\pm 0.2\,\%$ in the relevant operating range, and in the plant Kuwait plant with a vortex flow meter digital YEWFLO from Yokogawa [32] with an uncertainty of $\pm 0.75\,\%$. All temperatures were measured at each site using PT100 temperature sensors of class B, with regard to IEC 751/EN 60751, with an uncertainty of $\pm(0.30 + 0.00500\mathrm{K})$. The dataset for the evaluation covers half a year of operation in order to deliver profound insights. For the plants in Austria the measurements were taken from the beginning of July until the end of December 2018 and for the plant in Kuwait the data were taken from the beginning of September until the end of February 2019. The measurement data from the plants in Austria were recorded at a 1 s intervals and at a 6 min interval in Kuwait. The data from both sites were downsampled to an interval of 1 h to apply the method by calculating the mean for each measurement over hourly time bins.

3.2. Measures for validation

Since the adaptive solar forecasting method should be evaluated and compared for different solar heat producers having different nominal heat production rates, a scale-free measure is preferable for validation. The group of scale-free measures can be further split into measures based on *percentage errors* and measures based on the comparison of different methods (*relative measures*), which should both be further discussed starting with the percentage error.

Percentage errors are advantageous because they can be easily interpreted. However, the most commonly used percentage error, the Mean Absolute Percentage Error (MAPE), is unsuitable since it scales the forecast error based on the measured value, which can result in an infinite value if the solar heat production tends to zero. For this reason, other contributions in literature, e.g. [22,33,34], propose using the Mean Absolute Range Normalized Error (MARNE):

$$\mathrm{MARNE} = \frac{\mathrm{MAE}}{y_{max}} = \frac{\frac{1}{N_f}\sum_{i=1}^{N_f} |y_i - \hat{y}_i|}{y_{max}}, \tag{19}$$

i.e. the Mean Absolute Error (MAE) divided by the maximum value y_{max} of the variable to be forecast and N_f as the number of prediction values, whereby the nominal solar heat production of the plant can be used for the maximum value y_{max}. Since the MARNE is easy to interpret and cannot lead to infinite values, it is used to evaluate the forecasting method.

As a complement to that, the comparison to other (benchmark) methods is also illuminating. Relative measures are best suited to this purpose. In order to evaluate the superiority of the new method (*New*), it is compared to the seasonal naïve method (*SN*, Section 2.1) and the

Table 1
Table of the different solar collector systems investigated with their application, net collector area, orientation, nominal solar heat production and location.

Abbreviation	Application	Net collector area	Orientation	Nom. heat production	Location
$SF_{PH,S,S}$	Process heat	215 m^2	SE ($\gamma = -5°$)	150 kW	Kuwait
$SF_{HW,S,SW}$	Hot water	286 m^2	SW ($\gamma = 30°$)	172 kW	Austria
$SF_{HW,M,SE}$	Hot water	782 m^2	SE ($\gamma = -15°$)	469 kW	Austria
$SF_{HW,L,S}$	Hot water	2464 m^2	S ($\gamma = 0°$)	1478 kW	Austria
$SF_{HW,XL,mix}$	Hot water	3532 m^2	Mixed (–)	2119 kW	Austria

data sheet method (*DS*, Section 2.2). For this purpose the relative MAE is introduced

$$\frac{\mathrm{MARNE}_{\mathrm{New}}}{\mathrm{MARNE}_j} = \frac{\mathrm{MARNE}_{\mathrm{New}}/y_{\max}}{\mathrm{MARNE}_j/y_{\max}} = \frac{\mathrm{MAE}_{\mathrm{New}}}{\mathrm{MAE}_j} = \mathrm{RelMAE}_j\,, \tag{20}$$

with $j \in \{\mathrm{SN}, \mathrm{DS}\}$. For example, in case the $\mathrm{RelMAE}_{\mathrm{SN}}$ is greater than one, the naïve method performs better than the proposed one, and vice versa for a value smaller than one. All measures (MARNE, $\mathrm{RelMAE}_{\mathrm{SN}}$ and $\mathrm{RelMAE}_{\mathrm{DN}}$) are only evaluated for the hours between sunrise and sunset, since considering the whole day would erroneously lead to a lower error estimation. Before these measures are used for evaluation, the few parameters of the new method have to be determined.

3.3. Parameter determination

3.3.1. Forecast horizon N_{f}

The forecasting method should be used for EMS of large-scale solar thermal plants, repeatedly scheduling the production, storage and distribution of heat for the near future. This is typically done for a period of 24 h since the buffer storages of these plants are designed to store the energy of one perfectly sunny day; therefore, the forecast horizon is set to $N_{\mathrm{f}} = 24$ h.

3.3.2. Forecast of mean fluid temperature $\hat{\bar{T}}_{\mathrm{fl}}$

As described in Section 2.3, the forecast of the mean fluid temperature $\hat{\bar{T}}_{\mathrm{fl}}$ has to be set to a reasonable constant value to apply the method for forecasting the solar energy yield. It should be mentioned that it would be also possible to use a future course for $\hat{\bar{T}}_{\mathrm{fl}}$ or continuously adjust this will based on the planned operation to even increase the forecasting accuracy; however, a constant value already typically leads to good results and makes the method simpler to apply. This mean fluid temperature acts then as an offset on the predicted ambient temperature, influences the heat losses from the collector to the ambient and can be calculated from the operating conditions occurring at a plant. This was done separately for the two different applications considered, process heat and hot water, since they differ significantly regarding their temperature levels, resulting in two mean fluid temperatures: $\hat{\bar{T}}_{\mathrm{fl,PH}}$ for the process heat application and $\hat{\bar{T}}_{\mathrm{fl,HW}}$ for the hot water applications. For the process heat application this value was calculated by taking the average of the arithmetic mean temperature between the inlet and outlet over the whole measurement range when the plant ($SF_{PH,S,S}$) was in operation (the primary pump was switched on), resulting in $\hat{\bar{T}}_{\mathrm{fl,PH}} = 125\,°\mathrm{C}$. For the hot water applications, a mean temperature for each plant ($SF_{HW,S,S}$, $SF_{HW,M,SE}$, $SF_{HW,L,S}$, $SF_{HW,XL,mix}$) was first calculated as in the process heat case. In a next step, these values were averaged to obtain one value for the hot water applications, resulting in $\hat{\bar{T}}_{\mathrm{fl,HW}} = 73\,°\mathrm{C}$ in order to keep the number of parameters as small as possible.

3.3.3. Number of training days N_{d}

In contrast to the mean fluid temperature the optimal number of training days should be determined by applying the forecasting method separately to all solar heat producers from Table 1 to evaluate whether size and orientation have an impact on the number of training days N_{d}. Since the number of training days N_{d} only affects the prediction step of the method, the correction step (Section 2.3.3) could be neglected in this evaluation. In order to determine the optimal number of trainings days N_{d} the following procedure is carried out for each solar heat producer and each number of training days N_{d}, varying from 1 to 35 training days, whereby the forecast is calculated and evaluated hourly, including these 4 steps:

1. Updating the training data from the last N_{d} days of (measured) historical data, consisting of the solar yield, the mean fluid temperature, the global solar radiation and the ambient temperature.
2. Redetermining the model parameters β of the prediction model for the last hour.
3. Forecasting the solar heat production for a forecast horizon of 24 h, $N_{\mathrm{f}} = 24$ by using the measured global radiation and ambient temperature for the next 24 h.
4. Evaluating the forecast via the MARNE

These steps are repeated hourly, resulting in hourly updated forecasts for the solar heat production, each of which are valid for a forecasting horizon of 24 h and can be evaluated by the MARNE. Finally, the mean of these hourly calculated MARNEs is determined for each solar heat producer to find the optimal number of training days N_{d}.

The influence of different number of training days on the MARNE is shown exemplarily in Fig. 3, for a medium sized hot water solar plant $SF_{HW,M,SE}$. It can be seen that the MARNE increases at the beginning before falling rapidly. This means that using only one day for training (which is comparable to the naïve method) would perform better than using two or three training days, where similar was also observed for the other heat producers. Nevertheless, it can be also seen that the MARNE strongly decreases by about 40% when considering one week of training data as compared to only taking the day before for training. This shows that the training data for parameterization have a large impact on the forecasting accuracy. However, after a certain number of training days the value of the MARNE increases again, leading to the conclusion that the method cannot adapt itself appropriately if the training dataset is too large, making its performance worse. In Fig. 3,

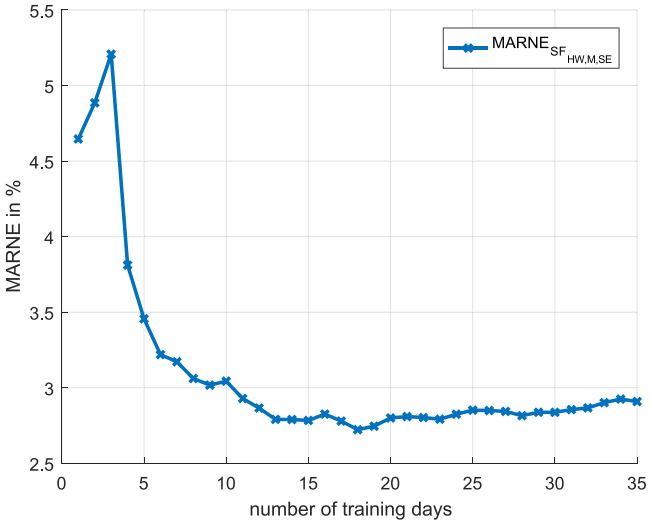

Fig. 3. Values of the Mean Absolute Range Normalized Error (MARNE) for different numbers of training days for an average sized heat producer $SF_{HW,M,SE}$.

Table 2
Optimal number of training days N_d for each heat producer based on the evaluation by the Mean Absolute Range Normalized Error (MARNE).

Abbreviation	Opt. number of training days N_d
$SF_{PH,S,S}$	20 d
$SF_{HW,S,SW}$	20 d
$SF_{HW,M,SE}$	18 d
$SF_{HW,L,S}$	18 d
$SF_{HW,XL,mix}$	18 d

it can be further seen that the MARNE values show a minimum for 18 training days, even if the curve is rather flat around this value. This was also observed for the other heat producers, i.e., that the MARNE values are rather similar around their minimum, with the optimal number of training days for each producer given in Table 2.

It can be concluded that, even for different climates and collector sizes, the data from the last two to three weeks should be considered. Taking fewer or more training data typically worsens the prediction. For the subsequent evaluation, the mean value over the optimal number of trainings days was chosen, leading to $N_d = 19$.

3.3.4. Correction gain c_c and correction horizon N_c

The correction uses the latest information for the forecast error to correct the following hours, which would otherwise only be considered 24 h later. The impact of the correction depends on the correction gain c_c, and how many forecast values are affected depends on the correction horizon N_c. As in the determination of the optimal number of training days N_t the optimal correction parameters (c_c, N_c) should be determined by applying the forecasting method to all solar heat producers listed in Table 1, but the correction step should be considered in the evaluation this time as well. Therefore all the parameters previously determined are used while varying the dimensionless correction gain c_c from 0 to 0.7 with a step size of 0.1 and varying the correction horizon N_c from 1 to 12 h. For each heat producer and each combination of c_c and N_c the forecasts have been calculated and evaluated hourly including the first three steps as described in Section 3.3.3 but additionally applying:

4. A correction for the prediction for a correction horizon of N_c hours by using a linearly decaying function starting from the current prediction error, and weighting it by the correction gain c_c.
5. Evaluating the forecast via the MARNE.

These steps are repeated hourly, resulting in hourly updated and corrected forecasts for the next 24 h for the solar energy yield. To determine the optimal training days, the accuracy of these forecasts is evaluated by calculating the MARNE for each forecast, yielding to hourly MARNEs. The minimum of the mean hourly MARNEs then determines the optimal values for c_c and N_c. Fig. 4 exemplarily shows the MARNE over the correction horizon N_c for different correction factors c_c for an average sized hot water heat producer $SF_{HW,M,SE}$. This figure shows that the value of the MARNE decreases as the correction gain c_c increases and that the correction horizon N_c reaches a certain point, where it increases again. In comparison to the determination of the optimal number of training days, the selection of the optimal correction parameters has little influence. However, on an overall scale, the correction leads to an improvement. Furthermore, it can be assumed that the influence of the correction is more important when applying real weather forecasts with higher error rates.

As in the determination of the optimal number of training days, it can be seen in Table 3 that the correction horizon N_c as well as the correction gain c_c are very similar for all heat producers, and no obvious correlation regarding the size of the heat producers can be detected. However, what can be seen in Fig. 4 and what was also

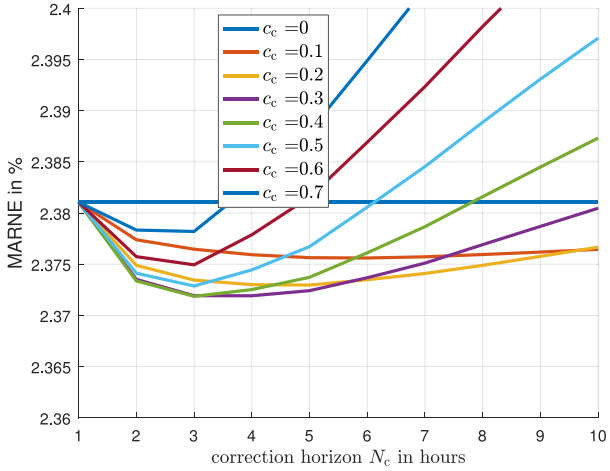

Fig. 4. Values of the Mean Absolute Range Normalized Error (MARNE) for different values of the correction gain c_c and the correction horizon N_c for an average sized heat producer $SF_{HW,M,SE}$.

Table 3
Optimal value of the correction gain c_c and the correction horizon N_c for each solar heat producer based on the evaluation by the Mean Absolute Range Normalized Error (MARNE).

Abbreviation	Opt.corr. gain c_c	Opt. corr. horizon N_c
$SF_{PH,S,S}$	0.40	3 h
$SF_{HW,S,SW}$	0.35	6 h
$SF_{HW,M,SE}$	0.30	4 h
$SF_{HW,L,S}$	0.45	5 h
$SF_{HW,XL,mix}$	0.45	5 h

experienced for the other fields is that in case the correction gain is large the correction horizon should be small and vice versa.

In conclusion, the mean value over all values of the correction gain and the correction horizon are used for the subsequent evaluation of the method, resulting in $c_c = 0.39$ and $N_c = 5$ h.

3.4. Validation results

After the parameters have been determined the performance of the new adaptive forecasting method should be comprehensively validated with measurement data, considering five important aspects:

1. *superiority over other forecasting methods* – by comparing it to two methods often used in literature: the seasonal naïve and the data sheet method.
2. *short-term performance* – by evaluating the performance for a representative week.
3. *long-term performance* – by evaluating the performance for six months of measurement data.
4. *wide applicability* – by applying it to a large-variety of different solar plants with different characteristics regarding application, size, orientation and climate conditions.
5. *real world applicability* – by using forecasts from a commercially available weather service provider to give insights into the performance of the method in practice.

To do so the section is split into three parts: First, in Section 3.4.1, a short-term evaluation is carried out for a representative week to evaluate the effects of changing ambient conditions, addressing aspects 1 and 2. Second, in Section 3.4.2, a long-term evaluation of the whole set of measurement data is carried out for all producers, assuming perfect weather forecasts (= measurements) and, addressing the aspects 1, 3 and 4. Third, in Section 3.4.3, the long-term evaluation is carried out again, this time using real weather forecasts from a commercially

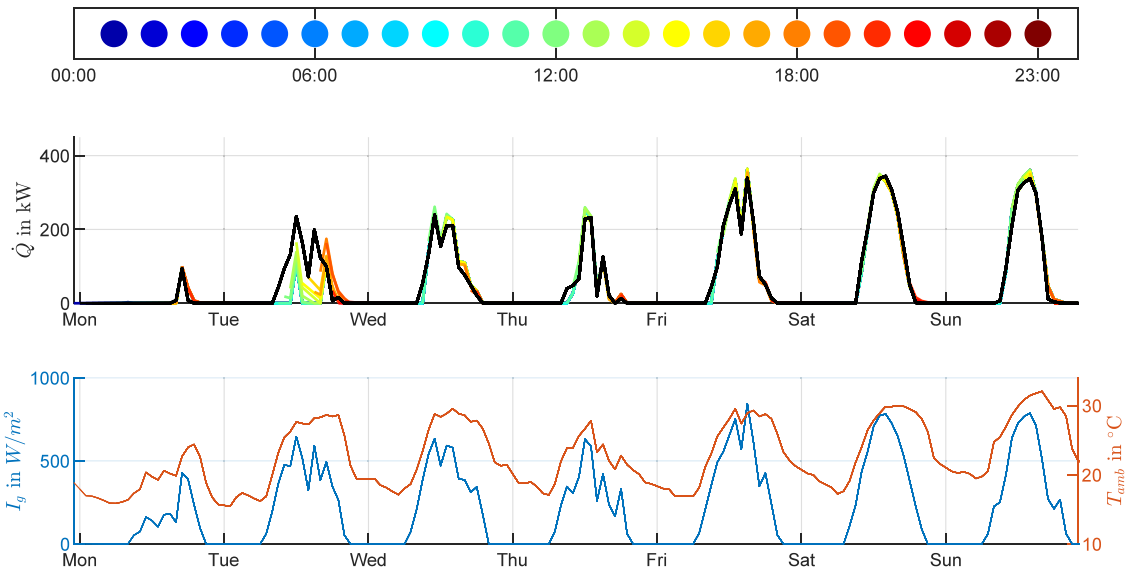

Fig. 5. Evaluation of the new forecasting method for a demanding week in July of a medium sized heat producer $SF_{HW,M,SE}$ in Austria, see Table 1. The upper graph shows the measured (black) and the forecast (colour) solar yield, with the colours indicating the hour of the day the respective forecast starts. The lower graphs shows the corresponding measured global radiation I_g and the measured ambient temperature T_{amb}.

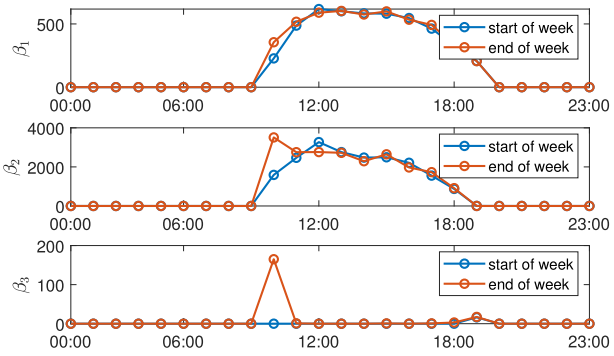

Fig. 6. Model parameters $\beta_1, \beta_2, \beta_3$ for each hour of the method at the start of a demanding week in July in comparison to their values at the end of the week changed by the automatic adaptation of the method.

weather service provider and additionally addressing aspect 5. Furthermore, in each section a comparison is made among the benchmark methods, the seasonal naïve and the data sheet method.

3.4.1. Short-term evaluation with perfect weather forecasts

In this section the forecast accuracy of the new method is evaluated for a representative week using a medium sized producer $SF_{HW,M,SE}$ listed in Table 1. Fig. 5 shows the measured solar heat production in black and the forecast in colour in the upper graph, with the colours indicating the hour of the day the respective forecast starts. The forecasts are calculated and evaluated hourly in the same way, as described in Section 3.3.4. In the lower graph, the measured values of the solar radiation I_g and the ambient temperature T_{amb} are shown, which are used to forecast the solar yield. Additionally, Fig. 6 shows the model parameters of the method $\beta_1, \beta_2, \beta_3$ for each hour at the start of the week in comparison to the end of the week which are continuously reparameterized, see Section 2.3.2.

From Fig. 5 it can be seen that for most of the days the proposed forecasting method shows very good results, allowing the prediction of the course of the produced solar heat with high accuracy, with only small deviations and also considering the right start and stop times of the system.

Sometimes, the application of the correction approach can be seen, for example on Tuesday: After the prediction shown at 14:00 in light green underestimates the prediction, the next predictions shown in yellow at 15:00, therefore, corrects the course upwards. After the second peak it can be seen that the correction makes the prediction even worse for a short time (dark orange at 19:00). However, as already discussed in the previous Section 2.3.3, the influence of the correction approach improves the total result but may worsen it for some time steps. From Fig. 6 it can be seen that the model parameters of some hours adapt over the week. For example, have the parameters β_1 and β_2 for the hours of 10:00 to 12:00 slightly increased their values as it is also the case for the parameter β_3 at 10:00. However, besides that, the parameters seem rather robust which has its origin in the relative large number of training days.

The overall MARNE evaluated for this week results in a MARNE of 4.15 % and the $RelMAE_{SN} = 0.31$ and $RelMAE_{DS} = 0.54$. This means that the method can be applied to forecast the solar heat production with a mean average deviation of 4.15 % of the nominal solar heat production, and is more than three times as accurate as the naïve forecast as well as nearly twice as accurate in comparison to the data sheet method. As described in Section 2.3.2, the optimal model parameters β are determined by solving an OLS problem by using the pseudo-inverse. When using an algorithm capable of solving the OLS problem while considering constraints, the MARNE for this week could be reduced to 3.72 %, improving the performance by about 25 %. However, to keep the method as general applicable as possible the pseudo-inverse was used for the further evaluations.

In this section the previously mentioned aspects are addressed: 1. *superiority over other forecasting methods* and 2. *short-term performance* were investigated, in the next section also 3. *long-term performance* and 4. *wide applicability*.

3.4.2. Long-term evaluation with perfect weather forecasts

After the short-term evaluation is made, the method is evaluated regarding its long-term performance and considering six months of data available. As in the short-term evaluation, this is done by considering a perfect weather forecast and, therefore, using measurements of the global solar radiation and the ambient temperature as inputs in order to evaluate the performance of the method itself independent of errors introduced by forecasts from a weather service provider. The long-term evaluation is made for each setting listed in Table 1 to prove

the wide-applicability and the results are compared to those of the two benchmarking methods: the seasonal naïve method (SN) and the data sheet method (DS), to demonstrate the superiority of the new method.

The first graph in Fig. 7 shows the resulting MARNEs for the new adaptive method (new) for each collector setting from Table 1 as well as in the second and third graphs the MARNE of the seasonal naïve method (SN) and the data sheet method (DS), respectively. In each of the graphs, a horizontal red dotted line illustrates the mean value of the respective MARNE over all collector settings. It should be noted that the MARNEs shown in Fig. 7 is for each heat producers the overall mean MARNE for the six months of measurement data when applying the forecast method at any hour of the year for predicting the solar yield for the next 24 h. The interested reader should be referred to the Appendix where the MARNE at each hour for forecasting the next 24 h are shown for each heat producer.

The wide applicability in terms of application, size, orientation and climate is given by comparing the MARNE in the first graph of Fig. 7 for each collector setting. It can be seen that size and orientation do not have a large influence on the forecast accuracy of the proposed method, resulting in similar MARNEs. A larger difference seems to exist regarding the application and climate, since the forecast of the plant $SF_{PH,S,S}$, which is used for process heat in Kuwait, suffers from a larger error. One reason for this can be that this plant has no large storage installed, since the heat is typically directly used by a process, making the usage of a constant mean absorber temperature a more critical issue since any variations in temperature coming from the process are not damped by a storage. The influence of climate is difficult to assess separately. However, it can be assumed that in the dry, relatively constant climate in Kuwait, it should be easier to make a forecast than it is in Central Europe. This can be also seen in the results of the seasonal naïve method which performs the best for $SF_{PH,S,S}$. From these results one can deduce that the consecutive days are similar. The opposite is true for the forecast using the data sheet parameters (lowest graph); here, it seems as though the smallest heat producers $SF_{HW,S,SW}$ can be forecast more poorly, while larger fields for hot water applications in Central Europe can be forecast much better. Overall, the two benchmarking methods perform similarly having a mean MARNE of about 7 % to 10 % while the proposed forecasting method has a MARNE of about only 3 %. This can even be more clearly shown by comparing the forecasting methods regarding the RelMAE, resulting in a $\text{RelMAE}_{SN} = 0.38$ and in a $\text{RelMAE}_{DS} = 0.28$. This shows the superiority of the method, since the proposed method is more than two times more accurate than the seasonal naïve method and more than three times more accurate than using the data sheet parameters. In this section, the previously mentioned aspects were investigated: 1. *superiority over other forecasting methods*, 3. *long-term performance* and 4. *wide applicability* and, in the last section, also 5. *real world applicability* is addressed.

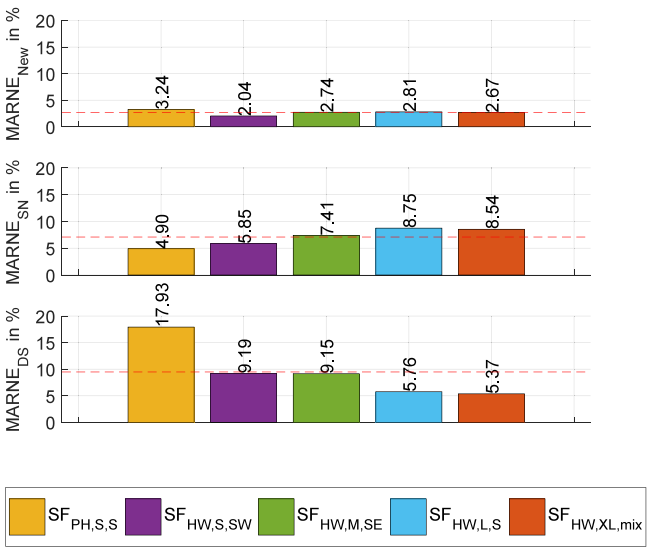

Fig. 7. Evaluation of the new adaptive method (New) for six months of measurement data using perfect weather forecast as inputs for five different heat producers listed in Table 1 in comparison to the seasonal naive method (SN) and the data sheet method (DS).

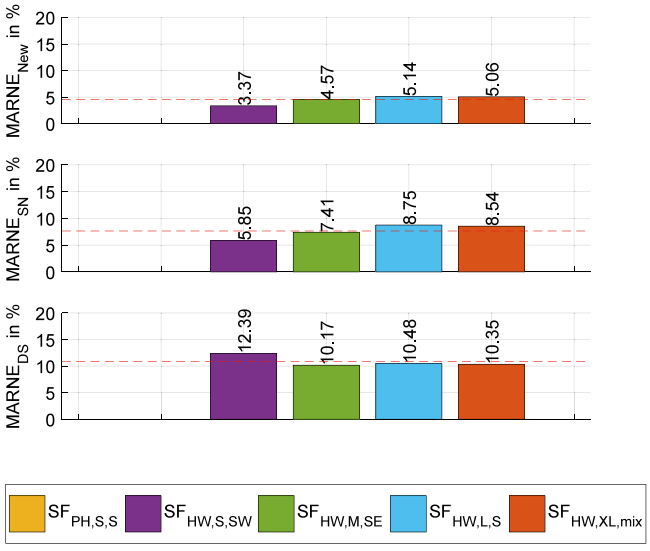

Fig. 8. Evaluation of the new adaptive method (New) for six months of measurement data using real weather forecast as inputs, for four different heat producers listed in Table 1 in comparison to the seasonal naive method (SN) and the data sheet method (DS).

3.4.3. Long-term evaluation with real weather forecasts

This long-term evaluation is done the same way as described in Section 3.4.2, but this time taking 'real' weather forecasts from a commercial weather service provider [27] in order to investigate the performance of the method in practice. Unfortunately, no weather forecast data for the region of Kuwait were available, so it is excluded from the evaluation. However, forecasting the weather conditions in such a dry region should be similar or even less challenging than forecasting conditions in middle Europe, leading at least to comparable results. Fig. 8 shows again the MARNEs for the different methods using 'real' weather forecasts as inputs.Since the seasonal naïve method does not consider forecasts of external factors, the results are the same as those seen in the previous section (see Fig. 7). Again, the red-dotted line illustrates the mean MARNE over all collector configurations for the specific method.

Regarding the wide applicability, the trend is the same as in the long-term evaluation using measurement data as forecasts but with a higher error rate. This rate increases relatively from the largest to the smallest field. Regarding the comparison between the performance of the new adaptive method to the benchmark methods, this leads to a $\text{RelMAE}_{SN} = 0.59$ for the seasonal naïve method and to a $\text{RelMAE}_{DS} = 0.42$ for the data sheet method. This means that even if the results are worse in comparison when taking real measurements the new method is still nearly twice as accurate as the seasonal naïve method and even more than twice as accurate as the data sheet method. In conclusion, the new method can be used to forecast the solar yield for a large variety of collector settings under real conditions with an average deviation of only 4.54 %, which can be seen as very promising to be used for energy management systems (EMS).

4. Conclusion and outlook

In this paper a new simple and adaptive forecasting method was presented for predicting the solar yield of large-scale flat-plate collector systems. In general, it combines the advantages of two commonly used forecasting methods, the seasonal naïve method and the data sheet method, to meet three important practical requirements: a simple implementation, an automatic adaption and a wide applicability.

Regarding the simple implementation, the method is straightforward with respect to its mathematical structure; it uses only linear models for each hour of the day, has just a few, easily tunable parameters and requires forecasts for only two external factors, the solar radiation and the ambient temperature, as inputs.

Regarding the automatic adaption, the method adapts self-adapts based on past measurement data by solving an overdetermined system of equations using the pseudo-inverse. Even if the pseudo-inverse can be easily implemented in any currently available controller, the investigations revealed that using a linear optimization algorithm capable of considering constraints led to better results and should be preferred if available.

Regarding the wide applicability, the investigations showed that the new method provides highly accurate forecasts with a Mean Absolute Range Normalized Error (MARNE) of about 5 % when using real weather forecasts as inputs, even when applied to a large variety of flat-plate collector systems. Furthermore, it outperforms the two most common forecasting methods, the seasonal naïve and the data sheet method, yielding results that are nearly twice as accurate.

Overall, the simple, adaptive forecasting method presented in this paper can be assumed to be highly suitable for use in energy management systems (EMS), helping to improve the performance and more efficiently integrate large-scale solar thermal plants into energy systems. The method allows to consider the expected solar yield in the operation of the energy system, which can reduce unnecessary operations of other heating systems (e.g., gas boiler) and improve the storage management in terms of the way in which the full solar yield can be stored.

As an outlook, one idea for future improvements is to apply the method to other collector technologies, for example evacuated tube collectors which are very prominent in China [1] or even concentrating collectors. While the former should be relatively easy, since evacuated tube collectors also rely on the European Standard EN12975 [25], the latter may require for a modification of the method to be made.

CRediT authorship contribution statement

Viktor Unterberger: Conceptualization, Formal analysis, Methodology, Writing - original draft, Software, Data curation, Project administration. **Klaus Lichtenegger:** Writing - review & editing, Supervision, Funding acquisition. **Valentin Kaisermayer:** Software, Visualization, Investigation. **Markus Gölles:** Conceptualization, Methodology, Funding acquisition, Supervision, Writing - review & editing. **Martin Horn:** Supervision, Writing - review & editing.

Declaration of competing interest

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

Acknowledgements

The research leading to these results has received funding from the Horizon 2020 program under Grant No. 792276, the Energieforschungsprogramm (engl.: Energy Research) under Grant No. 848776 and the COMET program under Grant No. 869341. The Energieforschungsprogramm is a research and technology program managed on behalf of the Austrian Ministry for Transport, Innovation and Technology (BMVIT) and the Austria Research Promotion Agency (FFG); the COMET program is also managed by the FFG and is co-financed by the Republic of Austria and the Federal Provinces of Vienna, Lower Austria and Styria.

Finally, the authors sincerely thank the partners from industry which helped to carry out this research: SOLID Solar Energy Systems GmbH and TVP Solar for providing the measurement data of their flat-plate collector systems and meteoblue AG for providing high quality weather forecasts.

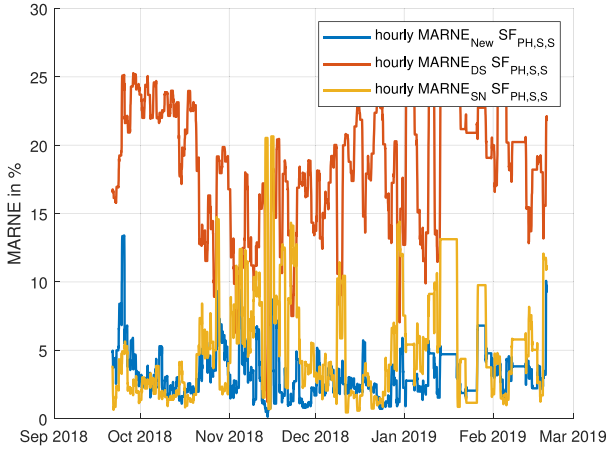

Fig. A.9. Values of the Mean Absolute Range Normalized Error (MARNE) evaluated at each hour of the measurement data for the heat producer SF$_{PH,S,S}$ listed in Table 1 for the new adaptive method (New) in comparison to the seasonal naïve method (SN) and the data sheet method (DS).

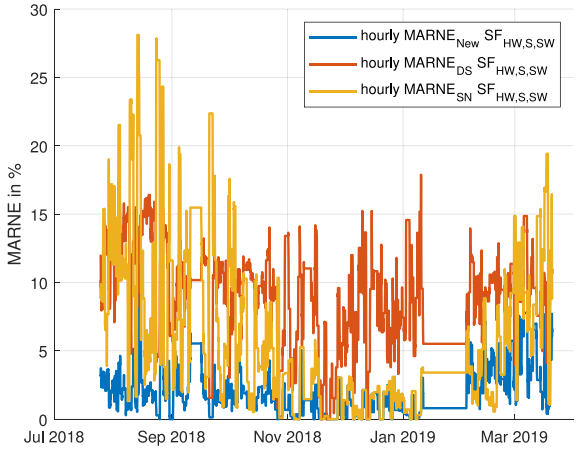

Fig. A.10. Values of the Mean Absolute Range Normalized Error (MARNE) evaluated at each hour of the measurement data for the heat producer SF$_{HW,S,SW}$ listed in Table 1 for the new adaptive method (New) in comparison to the seasonal naïve method (SN) and the data sheet method (DS).

Appendix. Hourly Mean Absolute Range Normalized Error (MARNE) for each heat producer using perfect weather forecasts

The Figs. A.9–A.13 show the MARNE for predicting the solar yield for the next 24 h evaluated at every hour of the measurement data considered, for each heat producer given in Table 1 for the new adaptive method (New) in comparison to the benchmarking methods, the seasonal naïve method (SN) and the data sheet method (DS).

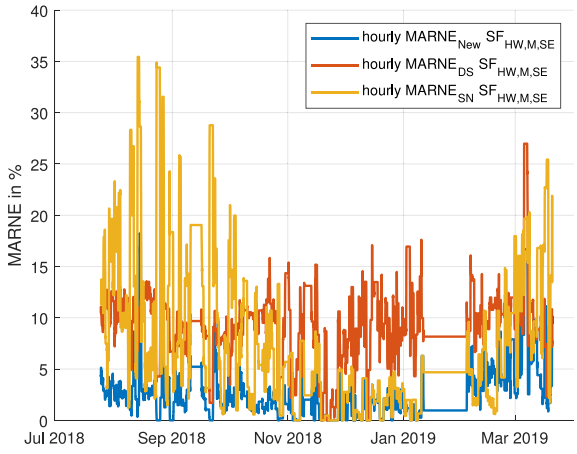

Fig. A.11. Values of the Mean Absolute Range Normalized Error (MARNE) evaluated at each hour of the measurement data for the heat producer $SF_{HW,M,SE}$ listed in Table 1 for the new adaptive method (New) in comparison to the seasonal naïve method (SN) and the data sheet method (DS).

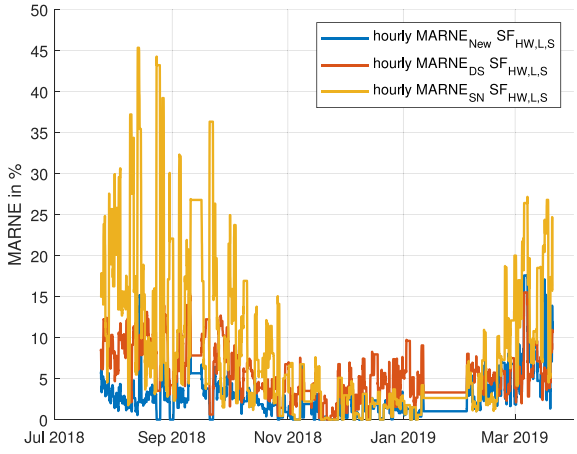

Fig. A.12. Values of the Mean Absolute Range Normalized Error (MARNE) evaluated at each hour of the measurement data for the heat producer $SF_{HW,L,S}$ listed in Table 1 for the new adaptive method (New) in comparison to the seasonal naïve method (SN) and the data sheet method (DS).

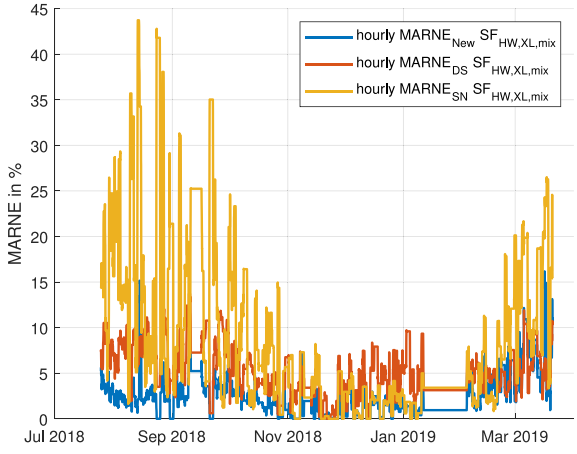

Fig. A.13. Values of the Mean Absolute Range Normalized Error (MARNE) evaluated at each hour of the measurement data for the heat producer $SF_{HW,XL,mix}$ listed in Table 1 for the new adaptive method (New) in comparison to the seasonal naïve method (SN) and the data sheet method (DS).

References

[1] Weiss W, Spörk-Dür M. Solar heat worldwide - global market development and trends in 2019. Resreport, IEA SHC Programme; 2020, URL: https://www.iea-shc.org/solar-heat-worldwide.

[2] Furbo S, Fan J, Perers B, Kong W, Trier D, From N. Testing, development and demonstration of large scale solar district heating systems. Energy Procedia 2015;70:568–73. http://dx.doi.org/10.1016/j.egypro.2015.02.162, International Conference on Solar Heating and Cooling for Buildings and Industry, SHC 2014.

[3] Moser A, Muschick D, Gölles M, Nageler P, Schranzhofer H, Mach T, et al. A MILP-based modular energy management system for urban multi-energy systems: Performance and sensitivity analysis. Appl Energy 2020;261:114342. http://dx.doi.org/10.1016/j.apenergy.2019.114342.

[4] Unterberger V, Lichtenegger K, Innerhofer P, Gerardts B, Gölles M. Evaluation of the potential for efficiency increase by the application of model-based control strategies in large-scale solar thermal plants. Int J Contemp Energy 2018;4(1):1–8.

[5] Dyreson A, Miller F. Night sky cooling for concentrating solar power plants. Appl Energy 2016;180:276–86. http://dx.doi.org/10.1016/j.apenergy.2016.07.118, URL: https://www.sciencedirect.com/science/article/pii/S0306261916310650.

[6] Das UK, Tey KS, Seyedmahmoudian M, Mekhilef S, Idris MYI, Deventer WV, et al. Forecasting of photovoltaic power generation and model optimization: A review. Renew Sustain Energy Rev 2018;81:912–28. http://dx.doi.org/10.1016/j.rser.2017.08.017.

[7] Van der Meer D, Widén J, Munkhammar J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. Renew Sustain Energy Rev 2018;81:1484–512. http://dx.doi.org/10.1016/j.rser.2017.05.212.

[8] Ahmad MW, Reynolds J, Rezgui Y. Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. J Cleaner Prod 2018;203:810–21. http://dx.doi.org/10.1016/j.jclepro.2018.08.207, URL: https://www.sciencedirect.com/science/article/pii/S0959652618325551.

[9] Pasamontes M, Álvarez J, Guzmán J, Berenguel M, Camacho E. Hybrid modeling of a solar-thermal heating facility. Sol Energy 2013;97:577–90. http://dx.doi.org/10.1016/j.solener.2013.09.024, URL: https://www.sciencedirect.com/science/article/pii/S0038092X13003861.

[10] Tian Z, Perers B, Furbo S, Fan J. Analysis and validation of a quasi-dynamic model for a solar collector field with flat plate collectors and parabolic trough collectors in series for district heating. Energy 2018;142:130–8. http://dx.doi.org/10.1016/j.energy.2017.09.135, URL: https://www.sciencedirect.com/science/article/pii/S0360544217316572.

[11] Bacher P, Madsen H, Perers B. Short-term solar collector power forecasting. In: Proceedings of ISES solar world conference 2011. 2011.

[12] Kalogirou S, Mathioulakis E, Belessiotis V. Artificial neural networks for the performance prediction of large solar systems. Renew Energy 2014;63:90–7. http://dx.doi.org/10.1016/j.renene.2013.08.049, URL: https://www.sciencedirect.com/science/article/pii/S0960148113004655.

[13] Kramer W, Bitterling M. Artificial neural networks (ANN) for the prediction of local outside temperatures and solar yields. In: Proceedings SWC2017. 2017, http://dx.doi.org/10.18086/swc.2017.22.0.

[14] Ghritlahre HK, Prasad RK. Application of ANN technique to predict the performance of solar collector systems - a review. Renew Sustain Energy Rev 2018;84:75–88. http://dx.doi.org/10.1016/j.rser.2018.01.001, URL: https://www.sciencedirect.com/science/article/pii/S1364032118300017.

[15] Khademi M, Jafarkazemi F, Moadel M, Razeghi A. Performance prediction of flat-plate solar collectors using MLP and ANFIS. J Basic Appl Sci Res 2013;3:196–200.

[16] Yaïci W, Entchev E. Adaptive neuro-fuzzy inference system modelling for performance prediction of solar thermal energy system. Renew Energy 2016;86:302–15. http://dx.doi.org/10.1016/j.renene.2015.08.028, URL: https://www.sciencedirect.com/science/article/pii/S0960148115302287.

[17] Correa-Jullian C, Cardemil JM, López Droguett E, Behzad M. Assessment of deep learning techniques for prognosis of solar thermal systems. Renew Energy 2020;145:2178–91. http://dx.doi.org/10.1016/j.renene.2019.07.100, URL: https://www.sciencedirect.com/science/article/pii/S0960148119311206.

[18] Kalogirou SA. Prediction of flat-plate collector performance parameters using artificial neural networks. Sol Energy 2006;80(3):248–59. http://dx.doi.org/10.1016/j.solener.2005.03.003.

[19] Xie H, Liu L, Ma F, Fan H. Performance prediction of solar collectors using artificial neural networks. In: 2009 international conference on artificial intelligence and computational intelligence, Vol. 2. 2009, p. 573–6. http://dx.doi.org/10.1109/AICI.2009.344.

[20] Fischer S, Frey P, Drück H. A comparison between state-of-the-art and neural network modelling of solar collectors. Sol Energy 2012;86(11):3268–77. http://dx.doi.org/10.1016/j.solener.2012.09.002, URL: https://www.sciencedirect.com/science/article/pii/S0038092X1200326X.

[21] Lee K, Heo J, Joo M, Lee S. Performance comparison for site-specific heat output prediction of solar collectors based on a modified collector efficiency equation model. Energy Procedia 2016;91:78–83. http://dx.doi.org/10.1016/j.egypro.2016.06.175, Proceedings of the 4th International Conference on Solar Heating and Cooling for Buildings and Industry (SHC 2015).

[22] Nigitz T, Gölles M. A generally applicable, simple and adaptive forecasting method for the short-term heat load of consumers. Appl Energy 2019;241:73–81. http://dx.doi.org/10.1016/j.apenergy.2019.03.012.

[23] Unterberger V, Nigitz T, Luzzu M, Muschick D, Gölles M. Adaptive methods for energy forecasting of production and demand of solar-assisted heating systems. In: Valenzuela O, Rojas F, Pomares H, Rojas I, editors. Theory and applications of time series analysis. Cham: Springer International Publishing; 2019, p. 287–99. http://dx.doi.org/10.1007/978-3-030-26036-1_20.

[24] Hyndman R, Athanasopoulos G. Forecasting: principles and practice. OTexts; 2018.

[25] Kovacs P. A guide to the standard EN 12975 - Deliverable D2.3. Technical report, SP – Technical Research Institute of Sweden, QAiST - IEE/08/593/SI2.529236; 2012.

[26] Duffie JA, Beckman WA. Solar engineering of thermal processes. 2nd ed.. Wiley New York; 1991, http://dx.doi.org/10.1002/9781118671603.

[27] meteoblue AG. Technical documentation. techreport, meteoblue AG; 2016, URL: https://www.meteoblue.com/.

[28] TYFOROP Chemie GmbH. TYFOCOR L konzentrat. Technical report, Tyforop Chemie GmbH; 2015, URL: https://www.tyfo.de/product/tyfocor-l/.

[29] Kipp & Zonen BV. SMP6 Pyranometer. 2016, URL: https://www.kippzonen.com/Product/358/SMP6-Pyranometer#.XuiXTdVCSHs.

[30] MTX Srl. MTX PCTRA056. 2018, URL: www.mtx.it/images/pdf/sensori/PCTRA056_PiranometroIClasse.pdf.

[31] KROHNE. Optiflux 4000. 2014, URL: https://krohne.com/en/products/flow-measurement/components-and-auxiliary-equipment-for-flow-measurement/flow-sensors/optiflux-4000/.

[32] YOKOGAWA. Yokogawa digital yewflo vortex flow meter. 2018, URL: https://www.yokogawa.com/solutions/products-platforms/field-instruments/flow-meters/vortex-flow-meters/digitalyewflo-dy/#Downloads__downloads_739.

[33] Potočnik P, Strmčnik E, Govekar E. Linear and neural network-based models for short-term heat load forecasting. Strojniski Vestn/J Mech Eng 2015;61:543–50. http://dx.doi.org/10.5545/sv-jme.2015.2548.

[34] Tratar LF, Strmčnik E. The comparison of Holt–Winters method and multiple regression method: A case study. Energy 2016;109:266–76. http://dx.doi.org/10.1016/j.energy.2016.04.115.