Renewable Energy 91 (2016) 224–232

Contents lists available at ScienceDirect

Renewable Energy

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

Improved multiple linear regression based models for solar collectors

Richárd Kicsiny

Department of Mathematics, Institute for Mathematics and Informatics, Szent István University, Páter K. u. 1., 2100 Gödöllő, Hungary

A R T I C L E I N F O

Article history:
Received 3 November 2015
Received in revised form
9 January 2016
Accepted 19 January 2016
Available online xxx

Keywords:
Solar collectors
Mathematical modelling
Black-box model
Multiple linear regression
Polynomial regression

A B S T R A C T

Mathematical modelling is the theoretically established tool to investigate and develop solar thermal collectors as environmentally friendly technological heat producers. In the present paper, the recent and accurate multiple linear regression (MLR) based collector model in Ref. [1] is empirically improved to minimize the modelling error. Two new, improved models called *IMLR model* and *MPR model* (where MPR is the abbreviation of multiple polynomial regression) are validated and compared with the former model (*MLR model*) based on measured data of a real collector field. The IMLR and the MPR models are significantly more precise while retaining simple usability and low computational demand. Many attempts to decrease the modelling error further show that the gained precision of the IMLR model cannot be significantly improved any more if the regression functions are linear in terms of the input variables. In the MPR model, some of the regression functions are nonlinear (polynomial) in terms of the input variables.

1. Introduction

It is highly important nowadays to study and develop solar thermal collectors within the framework of environmental protection. The theoretical tool for this purpose is mathematical modelling. Two main types of mathematical models exist for collectors: physically-based (or white-box) models describe exact and known physical laws, while black-box models represent some experienced or measured correlations empirically.

In the literature, there are numerous physically-based models. One of the most important and still often used ones is the Hottel-Whillier-Bliss model [2,3], which is also among the earliest models. The collector temperature is determined as a function of time and a space coordinate in this distributed model. Buzás et al. [4] proposed a simpler model based on the piston flow concept assuming that the collector temperature is homogeneous in space. This model is a linear ordinary differential equation (ODE) validated in Ref. [1] and is probably the simplest such (physically-based and ODE) model used in the practice (see e.g. Refs. [5–8]), which describes the transient processes of a collector with a proper accuracy.

The greatest advantage of black-box type models is that it is not needed to know precisely the physical laws of a collector in order to create an usable model. Nevertheless, the model may be rather precise even if it is mathematically simple as in the case of Ref. [1]. In the scope of solar thermal systems, the most widely used black-box model type may be the artificial neural network (ANN). In Ref. [9], the useful heat gained from a solar heating system as well as the temperature rise of the storage water were predicted with an ANN with an error of 7–10%, which is considered proper accuracy with respect to such systems [10]. The layer temperatures are modelled in a solar storage by means of an ANN elaborated in Ref. [11].

In particular, ANNs are frequently applied to model solar collectors separately as well. Generally speaking, ANNs are precise modelling approaches but quite troublesome to use because of the necessary training/learning process. A lot of measured data must be collected under various working conditions to train the ANN for gaining a satisfactory precision. E.g. in Ref. [12], the training requires the measured data of three months or in Ref. [13], the measurements of 17 days are needed to work out an ANN, which models a collector under similar circumstances as during the measurements. Only one so-called back-propagation algorithm (which is needed in the training) from the 11 available ones provides a proper approximation in the latter work, based on which, it can be easily concluded that the success of an ANN, and its computational demand, depend highly on the user's expertise. Also, the convergence of the algorithm indicating the end of the training can be time-consuming. Ref. [14] demonstrates the problem of uncertainty as well, where six separate ANNs are used to identify several collector parameters, so there is no general and exact instruction for designing appropriate ANNs. In fact, an

E-mail address: Kicsiny.Richard@gek.szie.hu.

http://dx.doi.org/10.1016/j.renene.2016.01.056
0960-1481/

Nomenclature

t	time, s
T_{out}	homogeneous temperature inside the collector (assumed to be the same as the outlet collector temperature), °C
I	global solar irradiance on the collector surface, W/m^2
T_a	ambient temperature of the collector, °C
T_{in}	inlet collector (fluid) temperature, °C
V	volume of the collector, m^3
v	(constant) flow rate inside the collector, m^3/s
τ_A	time delay before *Case A* or *A3*, s
τ_B	time delay before *Case B*, s
τ_1	time of flowing inside the collector from the inlet to the outlet when the pump is switched on permanently, s
τ_2	length of time between successive measurements on the collector, s

universal applicable algorithm, which ensures a reliable and fast design of a proper ANN for a collector is still missing based on Fischer at al [15].

Because of the above difficulties on complexity/uncertainty and time/computational demand, a precise and general but still simple black-box model has been proposed recently in Ref. [1], which can be applied fast and easily for many types of solar collectors. The model (called *MLR model* in short) is founded on standard methods from mathematical statistics, in particular, on multiple linear regression (MLR). Based on studies in the literature, the author found that MLR is a relatively rare black-box modelling tool in the field of solar collectors in spite of its simplicity. MLR can be applied to identify collector parameters in white-box models (see Refs. [16–18]) and to model collector efficiency in collector test methods [19,20]. In Ref. [21], the annual thermal performance of collectors is modelled as a function of the annual solar radiation by means of linear regression. In the mentioned MLR model, a typical day has been divided into sub-cases, for which, separate regression equations have been proposed to reach higher precision. The regression equations describe empirical relations immediately between input and output variables of collectors. Considering the high precision (with an error of 4.6%), simple usability and low computational demand of the MLR model, it is definitely worth improving further to maximize its precision while retaining simple usability and low computational demand. Accordingly, this improvement has been set as a future research task in the Conclusion of Ref. [1].

The following are the contributions of the present paper in details: the MLR model is empirically improved by means of inserting new operating sub-cases in a rather natural way. On the basis of measured data of a real collector, two new, improved models called *IMLR model* and *MPR model* (where MPR is the abbreviation of multiple polynomial regression) are identified (based on four days), validated (based on two months) and compared with the MLR model in view of accuracy. In the MPR model, some of the regression functions/equations are nonlinear (polynomial) in terms of the input variables. Polynomial regression is generally considered as a special linear regression, since the regression functions are linear in terms of the constant parameters (which are to be identified), although not linear in the input variables. If the regression functions are all linear, the regression can be called simple linear regression.

Matlab [22] has been used to carry out the needed calculations numerically. This software is widely used in the field of solar engineering to simulate different systems (see e.g. Ref. [23]).

The paper is structured as follows: Section 2 recalls the details of the recent MLR model for the Reader's convenience. In Section 3, the new IMLR and MPR models are worked out. These models are validated in Section 4 by means of measured and simulated data. Finally, conclusions and future research suggestions can be found in Section 5.

2. MLR model

For the Reader's convenience, the MLR model [1] is recalled in details in this section. Fig. 1 shows the studied solar collector.

The inputs in the MLR model are from appropriate T_{in}, I, T_a and T_{out} values. The output is from appropriate T_{out} values. The flow rate v is a prefixed positive constant or 0 according to the differential control, which is not only the most frequent control method but also the optimal or nearly optimal one many times [24,25].

Since the flow rate is bounded, only $T_{in}(t-\tau_1)$ can function as an input in the MLR model, where τ_1 is a (positive) time delay and the formed output is $T_{out}(t)$. Similarly, only former $I(t-\tau_2)$ and $T_a(t-\tau_2)$ values can function as inputs corresponding to the output $T_{out}(t)$ because these effects have bounded propagation speed. (For simplicity, the same delay (τ_2) is assumed for I and T_a.) Of course, an adequate former value of T_{out} affects $T_{out}(t)$ itself and functions essentially as the initial value in the MLR model at the time $(t-\tau_2)$. In the (black-box type) MLR model, distinct sub-models were identified for separate operating conditions. It is clear, for example, that the collector behaves very differently if the pump is off ($v=0$) or on ($v>0$) permanently. Under the same conditions, including a high enough solar irradiance, the collector temperature and thus T_{out} increases much more fast when the pump is off. The effect of T_{in} was neglected in permanently switched off case, since there is no flowing from the inlet to the outlet in the collector.

Assuming a typical day, when the increase of T_{out} is relatively high, three separate operating conditions were distinguished in accordance with Fig. 2.

The exact specification of each case is the following:

Case A: The pump is switched off permanently. *Case A* contains the term started at the beginning of the day and finished, when the pump is first switched on. All the terms, which begin at a time when the pump is permanently off for exactly τ_A time and finish at the next switch-on or at the end of the day, also belong to this case. (τ_A is the time, which is generally enough for T_{out} to become not fluctuating but permanently monotone, since, intentionally, frequent fluctuations are characteristics of *Cases C1* and *C2*.)

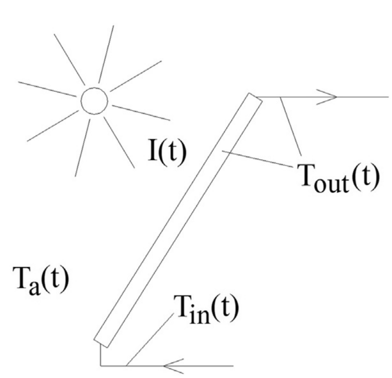

Fig. 1. The studied solar collector.