Mechanical Engineering Letters, Szent István University

Annual Technical-Scientific Journal of the Mechanical Engineering Faculty,
Szent István University, Gödöllő, Hungary

HU ISSN 2060-3789

Páter K. u. 1., Gödöllő, H-2103 Hungary
dekan@gek.szie.hu, www.gek.szie.hu,

Volume 14 (2016)

Multiple linear regression based models for solar collectors

Richárd KICSINY
Department of Mathematics,
Institute for Mathematics and Informatics

Abstract

Mathematical modelling is the theoretically established tool to investigate and develop solar thermal collectors as environmentally friendly technological heat producers. In the present survey, recent multiple linear regression (MLR) based collector models are presented and compared with one another and with a physically-based model, used successfully in many applications, by means of measured data. The MLR-based models, called *MLR model*, *SMLR model* and *IMLR model*, prove to be rather precise with a modelling error of 4.6%, 8.0% and 4.1%, respectively, which means that all MLR-based models are more or nearly the same accurate as the well-tried physically-based model. The SMLR model is the most, while the IMLR model is the least easy-to-apply MLR-based model with the lowest and the highest computational demand, respectively. Nevertheless, all MLR-based models have lower computational demand than the physically-based model. Accordingly, the MLR-based models are suggested for fast but accurate collector modelling.

Nomenclature

A: collector surface area, m^2;
c: specific heat capacity of the collector fluid, J/(kgK);
I: global solar irradiance on the collector surface, W/m^2;
t: time, s;
T_a: ambient temperature of the collector, °C;
T_{in}: inlet collector (fluid) temperature, °C;
T_{out}: outlet collector temperature (assumed to be the same as the homogeneous collector temperature in case of the physically-based model), °C;
U_L: overall heat loss coefficient of the collector, $W/(m^2K)$;
v: (constant) flow rate inside the collector, m^3/s;
V: volume of the collector, m^3;
η_0: optical efficiency of the collector, - ;
ρ: collector fluid density, kg/m^3;

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.

1. Introduction

Mathematical modelling is the most widely used and theoretically established tool to investigate and develop solar thermal collectors as environmentally friendly technological heat producers. The two main categories of mathematical models for collectors are physically-based models, which represent exact physical laws (based on theory), and black-box models, which describe empirical correlations (based on experiences or measurements).

Among the most important physically-based models, the Hottel-Whillier-Bliss model (Duffie and Beckman, 2006) may be the earliest, which is frequently used to date. This model determines the collector temperature as a function of time and space. Buzás et al. (1998) proposed a simpler model assuming that the collector temperature is homogeneous in space. This model is a linear ordinary differential equation (ODE) validated in (Kicsiny, 2014) and is likely the simplest physically-based model used in the practice (see e.g. (Kumar and Rosen, 2010; Buzás and Kicsiny, 2014)), but can still describe the transient collector processes with an appropriate accuracy. This model will be called *physically-based model* in short below.

The greatest advantage of black-box models is that it is not needed to know the physical laws of a collector precisely in order to create a model. Nevertheless, the model may be rather precise even if it is simple as in the case of (Kicsiny, 2014). The most frequent black-box model type is perhaps the artificial neural network (ANN) in the field of collectors. Generally, ANNs are accurate but rather troublesome to apply because of the so-called training process. The convergence of the algorithm indicating the end of a training session may be also time-consuming. According to Fischer at al. (2012), a conveniently usable algorithm ensuring a reliable and fast determination of an appropriate ANN for a collector is still missing.

Because of these problems, a simple and general but still accurate black-box model, which can be applied easily and fast for a wide range of solar collectors, has been recently worked out in (Kicsiny, 2014). The model is based on the well-known methods of mathematical statistics, more precisely, the multiple linear regression (MLR). Based on the literature, MLR is a rare black-box modelling technique in the field of collectors despite of its simplicity. Considering the high precision (with a modelling error of 4.6%), simple usability and low computational demand of the mentioned MLR-based model (*MLR model* in short), it was worth trying to simplify further the MLR model. Such a simplified model, called *SMLR model* (with a modelling error of 8.0%), has been worked out in (Kicsiny, 2015). On the other hand, it was also worth improving the MLR model to try to maximize the precision. Such an improved model, called *IMLR model* (with an error of 4.1%), has been worked out in (Kicsiny, 2016), where it has been empirically shown that the accuracy cannot be significantly improved any more if the regression equations are all linear in terms of the input variables.

In the present survey, as a summary of former works, the above MLR-based models are presented and compared with one another and with the physically-based model by means of measured data.

2. Physically-based and MLR-based models

For the Reader's convenience, the physically-based model, and the MLR-based models are recalled in this section. The scheme of the studied solar collector can be seen in Fig. 1.

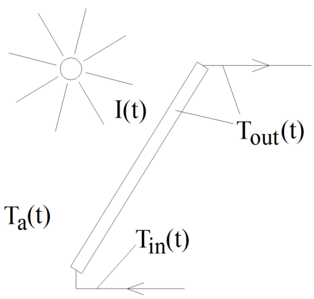

Figure 1. Scheme of the solar collector

2.1. Physically-based model

The physically-based model is the ODE of Eq. (1).

$$\frac{dT_{out}(t)}{dt}=\frac{A\eta_0}{\rho cV}I(t)+\frac{U_L A}{\rho cV}\left(T_a(t)-T_{out}(t)\right)+\frac{v}{V}\left(T_{in}(t)-T_{out}(t)\right) \tag{1}$$

2.2. MLR model

The inputs of the MLR model are from appropriately chosen values of T_{in}, I, T_a and T_{out}. The output is from appropriately chosen values of T_{out}. The flow rate value *v* is a fixed positive constant or 0.

Because of the boundedness of the flow rate, $T_{in}(t-\tau_1)$ can play a role as an input in the MLR model if $T_{out}(t)$ is the output, where the positive constant τ_1 is a time delay (more precisely, the time of flowing inside the collector from the inlet to the outlet when the pump is switched on permanently). Similar considerations hold for I and T_a as well because of the bounded propagation speed of their effects, so former $I(t-\tau_2)$ and $T_a(t-\tau_2)$ values can play roles as inputs in forming the output $T_{out}(t)$. (The time delays of I and T_a are assumed to be the same (τ_2) for the sake of simplicity.) Naturally, appropriate former value of T_{out} also affects the value of $T_{out}(t)$ and participates as the initial value of the MLR model at time $(t-\tau_2)$ in essence. Considering the collector as a black-box, distinct sub-models as parts of the MLR model have been identified for significantly different operating conditions. For example, the collector behaves different if the pump is on (v>0) or off (v=0) permanently. Even, the effect of T_{in} is neglected in permanently switched off case, since there is no flow between the collector inlet and outlet.

Considering a typical day, when the temperature increase of T_{out} is significant, three different, main operating cases are distinguished according to Fig. 2.

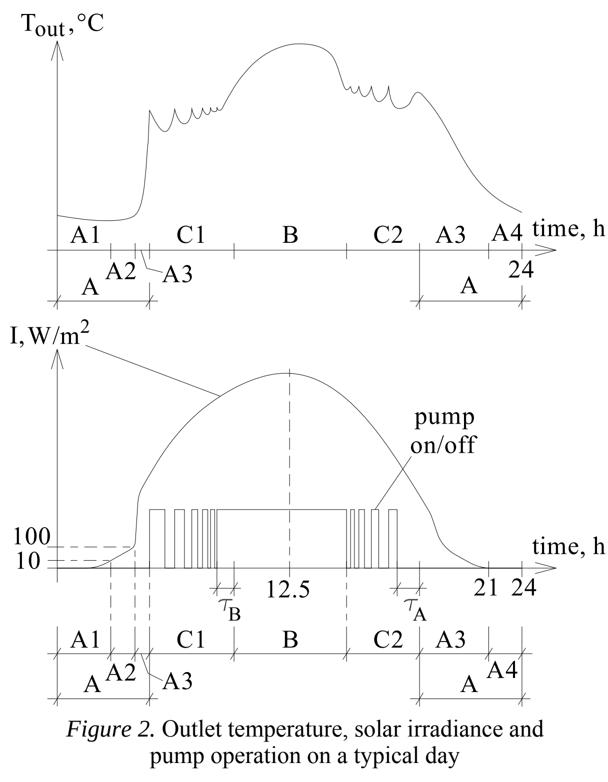

Figure 2. Outlet temperature, solar irradiance and pump operation on a typical day

Case A corresponds to permanently switched off pump, more precisely, 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.) Case B corresponds to permanently switched on pump, more precisely, Case B contains all the terms, which begin at a time when the pump is permanently on for exactly τ_B time and finish at the next switch-off. (τ_B is the time, which is generally enough for T_{out} to become not fluctuating but permanently monotone.) *Case C* corresponds to frequent switch-ons and -offs. It can be seen that there are two further significantly different operating cases within *Case C*: T_{out} basically increases before the solar noon and basically decreases after the solar noon, so *Case C* is divided into *Cases C1* and *C2*.

The MLR model is composed of the linear equations (2a)-(2d), which describe the corresponding sub-model of each operating case.

Case A:

$$T_{out,\mathrm{mod}}(t) = c_{I,A} I(t-\tau_2) + c_{a,A} T_a(t-\tau_2) + c_{out,A} T_{out}(t-\tau_2) + c_A \tag{2a}$$

Case B:

$$\begin{aligned} T_{out,\mathrm{mod}}(t) &= c_{in,B} T_{in}(t-\tau_1) + c_{I,B} I(t-\tau_2) + c_{a,B} T_a(t-\tau_2) + \\ &+ c_{out,B} T_{out}(t-\tau_2) + c_B \end{aligned} \tag{2b}$$

Case C1:

$$\begin{aligned} T_{out,\mathrm{mod}}(t) &= c_{in,C1} T_{in}(t-\tau_1) + c_{I,C1} I(t-\tau_2) + c_{a,C1} T_a(t-\tau_2) + \\ &+ c_{out,C1} T_{out}(t-\tau_2) + c_{C1} \end{aligned} \tag{2c}$$

Case C2:

$$\begin{aligned} T_{out,\mathrm{mod}}(t) &= c_{in,C2} T_{in}(t-\tau_1) + c_{I,C2} I(t-\tau_2) + c_{a,C2} T_a(t-\tau_2) + \\ &+ c_{out,C2} T_{out}(t-\tau_2) + c_{C2} \end{aligned} \tag{2d}$$

$c_{I,A}$, $c_{a,A}$, $c_{out,A}$, c_A, $c_{in,B}$, $c_{I,B}$, $c_{a,B}$, $c_{out,B}$, c_B, $c_{in,C1}$, $c_{I,C1}$, $c_{a,C1}$, $c_{out,C1}$, c_{C1}, $c_{in,C2}$, $c_{I,C2}$, $c_{a,C2}$, $c_{out,C2}$, c_{C2} are constant parameters. According to the definition of τ_2, the measurements take place at times $t = 0, \tau_2, 2\tau_2, 3\tau_2, \ldots$ The modelled value of T_{out} (that is $T_{out,\mathrm{mod}}$) is determined at times $t = \tau_2, 2\tau_2, 3\tau_2, \ldots$

from the measured values of $I(t-\tau_2)$, $T_a(t-\tau_2)$, $T_{out}(t-\tau_2)$ and $T_{in}(t-\tau_1)$ based on Eqs. (2a)-(2d). See (Kicsiny, 2014) for more details.

2.3. SMLR model

The MLR model is simplified to the SMLR model in the way of merging *Cases A*, *B*, *C1* and *C2*. Thus there is only one operating case with one mathematical relation (see Eq. (3)) here.

$$T_{out,\mathrm{mod}}(t)=c_{in}T_{in}(t-\tau_1)+c_I I(t-\tau_2)+c_a T_a(t-\tau_2)+c_{out}T_{out}(t-\tau_2)+c \qquad (3)$$

c_{in}, c_I, c_a, c_{out} and c are constant parameters. See (Kicsiny, 2015) for more details.

2.4. IMLR model

The IMLR model is similar to the MLR model. The main differences are the following (see also Fig. 2):

1. The (largest) operating case *Case A* is divided into four sub-cases, *Cases A1*, *A2*, *A3* and *A4*, as follows: *Case A1* consists of the time period from the beginning of the day to the time when the solar irradiance is first greater than 10 W/m^2. This case practically belongs to the term of no irradiance in the first half of the day. *Case A2* consists of the time period from the end of *Case A1* to the time when the solar irradiance is first greater than 100 W/m^2. Usually, this time is followed by a very intensive increase in the irradiance, so this is apparently the time of sunrise, when the irradiance changes from (mostly) diffuse to (mostly) direct. *Case A3* consists of the time periods besides *Cases A1*, *A2* and *A4* (see below) within *Case A*. *Case A4* contains the last three hours of the day. In essence, the term after *Case C2* corresponds to the free cooling of the collector from a relatively high temperature. Based on experiments, this section cannot be modelled well with a single relation, so it should be divided into sub-sections. Empirically, the problem can be solved well with only two sub-parts if the last three hours are separated.

2. The coefficients of the zeroth-order members (cf. c_A, c_B, c_{C1} and c_{C2} in Eqs. (2a)-(2d) in the MLR model) are set zero in Eqs. (4a)-(4g) below. This set is in line with the physical phenomenon that the collector (outlet) temperature must be zero if all the inputs T_{in}, I, T_a and the previous collector temperature are zero. Based on experiments, this natural constraint results in a bit lower precision in the identification but higher precision in the validation, that is the modelling error decreases.

The linear equations (4a)-(4g) correspond to the operating cases of the IMLR model.

Case A1:

$$T_{out,\mathrm{mod}}(t)=c_{a,A1}T_a(t-\tau_2)+c_{out,A1}T_{out}(t-\tau_2) \qquad (4a)$$

Case A2:

$$T_{out,\mathrm{mod}}(t)=c_{I,A2}I(t-\tau_2)+c_{a,A2}T_a(t-\tau_2)+c_{out,A2}T_{out}(t-\tau_2) \tag{4b}$$

Case A3:

$$T_{out,\mathrm{mod}}(t)=c_{I,A3}I(t-\tau_2)+c_{a,A3}T_a(t-\tau_2)+c_{out,A3}T_{out}(t-\tau_2) \tag{4c}$$

Case A4:

$$T_{out,\mathrm{mod}}(t)=c_{a,A4}T_a(t-\tau_2)+c_{out,A4}T_{out}(t-\tau_2) \tag{4d}$$

Case B:

$$\begin{aligned}T_{out,\mathrm{mod}}(t)&=c_{in,B}T_{in}(t-\tau_1)+c_{I,B}I(t-\tau_2)+c_{a,B}T_a(t-\tau_2)+\\&+c_{out,B}T_{out}(t-\tau_2)\end{aligned} \tag{4e}$$

Case C1:

$$\begin{aligned}T_{out,\mathrm{mod}}(t)&=c_{in,C1}T_{in}(t-\tau_1)+c_{I,C1}I(t-\tau_2)+c_{a,C1}T_a(t-\tau_2)+\\&+c_{out,C1}T_{out}(t-\tau_2)\end{aligned} \tag{4f}$$

Case C2:

$$\begin{aligned}T_{out,\mathrm{mod}}(t)&=c_{in,C2}T_{in}(t-\tau_1)+c_{I,C2}I(t-\tau_2)+c_{a,C2}T_a(t-\tau_2)+\\&+c_{out,C2}T_{out}(t-\tau_2)\end{aligned} \tag{4g}$$

$c_{a,A1}$, $c_{out,A1}$, $c_{I,A2}$, $c_{a,A2}$, $c_{out,A2}$, $c_{I,A3}$, $c_{a,A3}$, $c_{out,A3}$, $c_{a,A4}$, $c_{out,A4}$, $c_{in,B}$, $c_{I,B}$, $c_{a,B}$, $c_{out,B}$, $c_{in,C1}$, $c_{I,C1}$, $c_{a,C1}$, $c_{out,C1}$, $c_{in,C2}$, $c_{I,C2}$, $c_{a,C2}$, $c_{out,C2}$ are constant parameters. τ_1 and τ_2 are the same as in the MLR model. See (Kicsiny, 2016) for more details.

3. Comparison

The below results of the models used in this comparison are from (Kicsiny, 2014; 2015 and 2016). The identification and the validation of the models are based on the same days. The used real flat plate collector field of 33.3 m^2 (Farkas et al., 2000) at the Szent István University (SZIU) in Gödöllő, Hungary (*SZIU collector* in short) is also the same. T_{out}, T_{in}, I, T_a and v are measured once in every minute at the SZIU collector. The measured value of T_{out} serves only for identification and comparison purposes, the measured value $T_{out}(0)$ is fed into the models as initial condition.

Identification

Four measured days (2nd July 2012, 24th June 2012, 28th June 2012 and 8th June 2012) have been selected for the identification in such a way that they cover a wide range of possible operating conditions of a selected season (summer). The constant U_L is identified in case of the physically-based model in such a way that the time average of the absolute difference between the modelled and measured outlet temperatures that is the average of absolute error is minimal with respect to the whole identification period. The constants $c_{I,A}$, $c_{a,A}$, $c_{out,A}$, c_A, $c_{in,B}$, $c_{I,B}$, $c_{a,B}$, $c_{out,B}$, c_B, $c_{in,C1}$, $c_{I,C1}$, $c_{a,C1}$, $c_{out,C1}$, c_{C1}, $c_{in,C2}$, $c_{I,C2}$, $c_{a,C2}$, $c_{out,C2}$, c_{C2} are identified in the MLR model, the constants c_{in}, c_I, c_a, c_{out}, c are identified in the SMLR model and the constants $c_{a,A1}$, $c_{out,A1}$, $c_{I,A2}$, $c_{a,A2}$, $c_{out,A2}$, $c_{I,A3}$, $c_{a,A3}$, $c_{out,A3}$, $c_{a,A4}$, $c_{out,A4}$, $c_{in,B}$, $c_{I,B}$, $c_{a,B}$, $c_{out,B}$, $c_{in,C1}$, $c_{I,C1}$, $c_{a,C1}$, $c_{out,C1}$, $c_{in,C2}$, $c_{I,C2}$, $c_{a,C2}$, $c_{out,C2}$ are identified in the IMLR model. Independent standard MLR routines have been applied based on the measured data of each separate operating case for the identification of the three MLR-based models. The standard MLR routine (based on least squares method) is well-known and available in most statistical and spreadsheet programs (SPSS, Excel, etc.).

Table 1 contains the average of error (time average of the difference between the modelled and measured outlet temperatures) and the average of absolute error (time average of the absolute difference between the modelled and measured outlet temperatures) values for two days (2nd July 2012, 28th June 2012) of the identification of all models. The average of absolute error values are presented in proportion to the difference between the daily maximal and minimal measured outlet temperature values as well, in %. The mean of these values with respect to all of the four days of the identification is also presented in Table 1 (7.8 % for the physically-based, 4.7 % for the MLR, 6.6 % for the SMLR and 3.2 % for the IMLR model).

Validation

In the validation, all identified models are applied with the corresponding measured inputs of the remaining two summer months. The outlet temperature is modelled in the validation (not measured, as in the identification). The modelled days are from 3rd July to 31st August 2012, which are 56 days according to minor technical interruptions.

Table 1 contains the average of error and the average of absolute error values for two days (3rd August 2012, 5th August 2012) of the validation of all models. The average of absolute error values are presented in proportion to the difference between the daily maximal and minimal measured outlet temperature values as well, in %. The mean of these values with respect to all of the 56 days of the validation is also presented in Table 1 (7.8 % for the physically-based, 4.6 % for the MLR, 8.0 % for the SMLR and 4.1 % for the IMLR model).

Table 1. Average of error and average of absolute error values with the models

			Physically-based model	MLR model	SMLR model	IMLR model
Identification	2nd July (smooth operation)	Average of error	-1.86 °C	-0.47 °C	1.43 °C	-0.53 °C
		Average of absolute error	4.33 °C; 7.0%	2.79 °C; 4.6%	3.88 °C; 6.3%	1.64 °C; 2.7%
	28th June (intermittent operation)	Average of error	-1.26 °C	-0.23 °C	-2.87 °C	-0.80 °C
		Average of absolute error	4.35 °C; 7.5%	3.01 °C; 5.2%	3.39 °C; 5.8%	1.84 °C; 3.2%
	Mean % value for the whole identification (four days)	Average of absolute error	7.8%	4.7%	6.6%	3.2%
Validation	3rd August (smooth operation)	Average of error	-1.38 °C	-1.31 °C	-0.12 °C	-1.18 °C
		Average of absolute error	4.70 °C; 7.4%	2.85 °C; 4.5%	3.71 °C; 5.8%	1.92 °C; 3.0%
	5th August (intermittent operation)	Average of error	-2.57 °C	-1.58 °C	-1.00 °C	-1.90 °C
		Average of absolute error	4.66 °C; 8.0%	3.07 °C; 5.2%	3.95 °C; 6.7%	2.30 °C; 3.9%
	Mean % value for the whole validation (3rd July – 31st August)	Average of absolute error	7.8%	4.6%	8.0%	4.1%

As examples of the comparison, Figs. 3, 4 show the modelled and measured outlet temperatures in case of the physically-based and MLR models and in case of the MLR and IMLR models, respectively, for the same day of the validation. The operating state (on/off) of the pump is also shown in the figures.

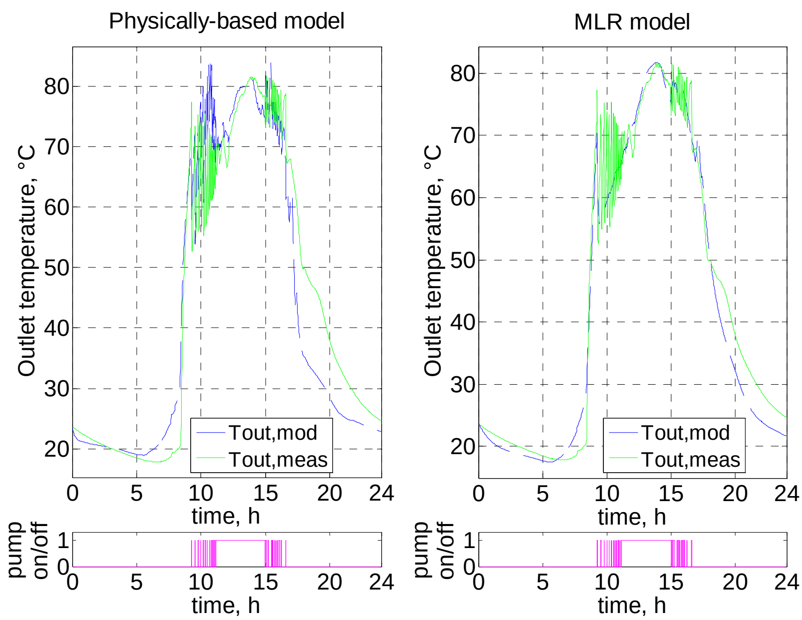

Figure 3. Modelled $T_{out,\mathrm{mod}}$ and measured $T_{out,meas}$ collector temperatures on 3rd August 2012 in case of the physically-based and MLR models

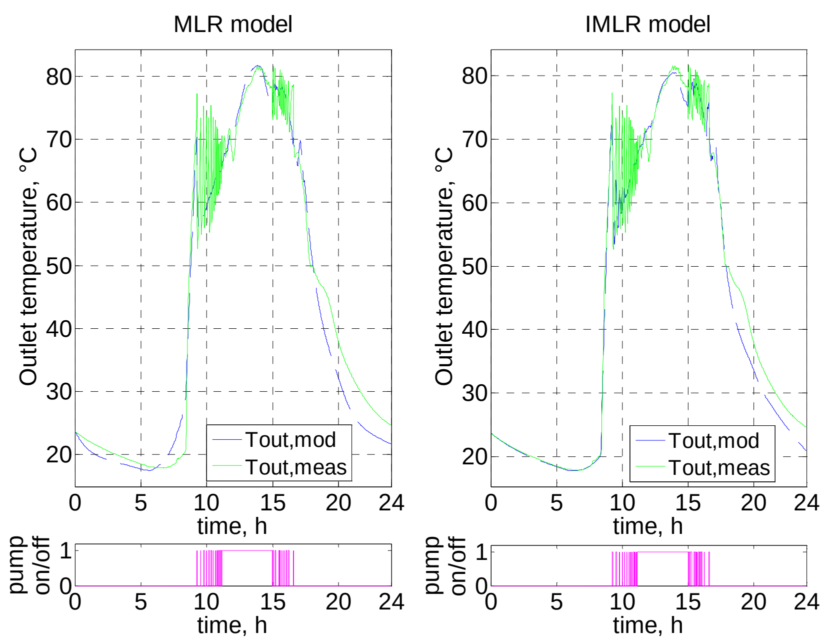

Figure 4. Modelled $T_{out,\mathrm{mod}}$ and measured $T_{out,meas}$ collector temperatures on 3rd August 2012 in case of the MLR and IMLR models

Based on Eqs. (1), (2a)-(2d), (3) and (4a)-(4g), it is not difficult to see that the SMLR model is the most, while the IMLR model is the least easy-to-apply MLR model with the lowest and the highest computational demand, respectively. Nevertheless, all MLR-based models with simple linear algebraic equations have lower computational demand than the physically-based model realized with an ODE.

Conclusion

In the present survey, three recent MLR-based models, called *MLR model*, *SMLR model* and *IMLR model*, have been presented and compared with one another and with a well-tried physically-based model by means of measured data. According to the results, all the MLR, SMLR and IMLR models have proved to be rather precise with a modelling error of 4.6%, 8.0% and 4.1%, respectively, which means that all MLR-based models are more or nearly the same accurate as the physically-based model with an error of 7.8%.

The SMLR model is the most, while the IMLR model is the least easy-to-apply MLR model with the lowest and the highest computational demand, respectively. Nevertheless, all the MLR-based models applying simple linear algebraic equations have lower computational demand than the physically-based model realized with an ODE. Accordingly, the MLR-based models can be suggested for fast but accurate collector modelling.

Acknowledgement

The author thanks his colleagues in the Department of Mathematics in the Faculty of Mechanical Engineering (Szent István University) for their support.

This paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

References

[1] Buzás, J., Farkas, I., Biró, A., Németh, R. (1998), Modelling and simulation of a solar thermal system, Mathematics and Computers in Simulation, Vol. 48, pp. 33-46.

[2] Buzás, J., Kicsiny, R. (2014), Transfer functions of solar collectors for dynamical analysis and control design, Renewable Energy, Vol. 68, pp. 146-155.

[3] Duffie, J.A., Beckman, W.A. (2006), Solar engineering of thermal processes, 3rd ed., John Wiley and Sons, New York.

[4] Farkas, I., Buzás, J., Lágymányosi, A., Kalmár, I., Kaboldy, E., Nagy, L. (2000), A combined solar hot water system for the use of swimming pool and

kindergarten operation, Energy and the environment, Vol. I. /ed. by B. Frankovic/, Croatian Solar Energy Association, Opatija, 2000., pp. 81-88.
[5] Fischer, S., Frey, P., Drück, H. (2012), A comparison between state-of-the-art and neural network modelling of solar collectors, Solar Energy, Vol. 86, pp. 3268-3277.
[6] Kicsiny, R. (2014), Multiple linear regression based model for solar collectors, Solar Energy, Vol. 110, pp. 496-506.
[7] Kicsiny, R. (2015), Simplified multiple linear regression based model for solar collectors, Hungarian Agricultural Engineering, Vol. 28, pp. 11-14.
[8] Kicsiny, R. (2016), Improved multiple linear regression based models for solar collectors, Renewable Energy, Vol. 91, pp. 224-232.
[9] Kumar, R., Rosen, M.A. (2010), Thermal performance of integrated collector storage solar water heater with corrugated absorber surface, Applied Thermal Engineering, Vol. 30, pp. 1764-1768.