Black-box model for solar storage tanks based on multiple linear regression

Richárd Kicsiny

Department of Mathematics, Institute of Environmental Systems, Faculty of Mechanical Engineering, Szent István University, Páter K. u. 1., 2100 Gödöllő, Hungary

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

Tel.: +36 28522000/1413, fax: +3628410804

Abstract

Developing easy-to-use mathematical models for describing temperatures of solar storage tanks is important for the practice, since solar storages are unavoidable elements in solar heating systems, where some heat should be stored in the form of hot fluid. In this paper, a new, general and easy-to-apply black-box model, called LR model (where LR is the abbreviation of linear regression), is proposed for solar storages on the basis of multiple linear regression. This linear model may be the simplest black-box type model, which can follow the transient processes of solar storages precisely. Accordingly, the LR model proves to be more precise than a slightly modified version of a physically-based storage model used successfully for different applications in the literature. The modified physically-based model accounts for the short circuit effect occurring in storages. Comparing measured and simulated data on a real solar storage, both models are validated and their efficiency is discussed in details. The general and simple usability of the LR model is mentioned and future research proposals are given.

Keywords:
Solar storage tanks; Modelling; Black-box; Linear regression; On/off operation

Nomenclature

t: time, s;

Time-dependent variables

T_{cold}: inlet temperature of the cold fluid to be heated in the storage, °C;

T_e: temperature of the environment of the storage, °C;

T_{in}: inlet temperature to the storage from the heating loop, °C;

T_{load}: temperature of the outlet fluid discharged from the storage by the consumer, °C;

T_s: geometrical average temperature inside the storage, °C;

$T_{s,meas}$: measured geometrical average temperature inside the storage, °C;

$T_{s,\mathrm{mod}}$: modelled geometrical average temperature inside the storage, °C;

T_{out}: outlet temperature from the storage to the heating loop, °C;

v: pump flow rate in the heating loop (according to on/off operation), m^3s^{-1};

v_{load}: flow rate of the consumption load, m^3s^{-1}

Constant parameters

A: outside area of the storage, m^2;

c: specific heat capacity of the fluid in the storage, $Jkg^{-1}K^{-1}$;

c_v : modifying coefficient related to the short circuit of the heating loop inside the storage, -;
k : heat loss coefficient of the storage to the environment, $Wm^{-2}K^{-1}$;
V : volume of the storage, m^3;
ρ : mass density of the fluid in the storage, kgm^{-3};
Δt : time period between successive measurements, s;
τ : time delay with respect to the effects of the inlets to (the interior of) the storage, s;
τ_A : time lag before *Case A* in the LR model, s;
τ_B : time lag before *Case B* in the LR model, s

Abbreviations

ANN: artificial neural network;
CFD: computational fluid dynamics;
LR: linear regression;
MLR: multiple linear regression;
ODE: ordinary differential equation;
PDE: partial differential equation

1. Introduction

Developing mathematical models for describing temperatures of solar storage tanks is of great importance for the practice, since these elements are unavoidable in any solar heating systems, where some heat should be stored in the form of hot fluid. There are two chief kinds of mathematical models of hydraulic storages: physically-based models (or white-box models), describing exact physical relations (on the basis of theory) and black-box type models, standing for empirical correlations based on (measured) experiences. Grey-box type models are mixed versions of physically-based and black-box ones.

The literature contains many physically-based models. In [1], both a mixed storage model of one dimension (assuming homogeneous temperature) and a multidimensional model of several nodes for stratified storages are presented. In the first one-node model, an ordinary differential equation (ODE), while in the latter one, a system of ODEs represent the corresponding mathematical description. Zeghib and Chaker [2] propose a similar model, where the storage is divided into several layers (nodes) of homogeneous temperatures, each of which is represented with a single ODE.

If only one node is applied, we recall in essence the linear ODE model from Buzás and Farkas [3]. This is one of the simplest physically-based ODE models, which can follow the transient processes of solar storages with proper accuracy. This is confirmed by many works showing also the successful and simple usability of the model (see e.g. [4]). Simple usability is an important advantage featuring linear models. In this paper, the slightly modified version of this ODE is used, which is completed with a modifying coefficient c_v responsible for the so-called short circuit effect. The short circuit effect is the following: although the inlet heating flow rate v can be measured, its effective value may be lower because of that a part of the corresponding fluid stream flowing into the storage (from the heating loop) may directly turn back and leave the storage (to the same heating loop) through the outlet pipe without any heating of the bulk of the storage fluid. Based on experiences, this effect can be significant and should be considered for better modelling precision. In particular, the short circuit effect influences the thermal efficiency of storages [5] and, naturally, extends the stagnation time of the fluid [6], which may be unhealthy in drinking water applications.

Temperature stratified type, heat balance, turbulent mixing and displacing mixing storage models on different physical phenomena and mathematical relations with both time and space dependence are summarized in [7]. The first two models are distributed ones described by

partial differential equations (PDEs) while the latter two ones are discretized models. A more recent, dimensionless PDE model for storages of solar power plants can be found in [8].

CFD (computational fluid dynamics) codes for computers are important tools of realistic modelling of storages [9, 10, 11, 12] using many physical equations, solving them with numerical methods, but, the multistep application process of them requires expert users, furthermore, their computational demand are high. In [13], the measured data and the results simulated with a CFD software are compared (see Fig. 6 in [13]), based on which the modelling error (the average absolute difference of the measured and simulated temperatures divided by the (positive) difference between the maximal and minimal measured values) is about 30% at the top of the storage and 10% in the middle. Another frequent modelling code is TRNSYS, working also with numerical solution methods, which can deal with stratification, internal auxiliary heating, inner heat conduction and mixing with respect to storages [14].

The difficulties of the exact physically-based modelling caused by the complex/complicated physical phenomena, like the effect of the inlet fluid flow on the thermal stratification [12, 15], can be often overcome by means of empirical black-box type models. The most frequent black-box model version may be the artificial neural network (ANN) for thermal engineering applications. Kalogirou et al. [16] model the temperature change in a solar storage with an ANN with a modelling error of 7-10%. This can be called proper precision for similar systems [17]. An ANN is established in [18] to predict the temperatures of the layers in a storage tank for domestic water. In general, ANNs are precise tools for modelling but quite arduous to use because of the so-called, quite complex and time consuming, training/learning process. The uncertainty problem that there are no general and detailed algorithms for designing successful ANNs, is also significant. Similar modelling tools, for example genetic algorithms, can be found as well [19]. Another black-box method is given in [20], where a general linear storage model is used describing the simple sum of the unperturbed process and a perturbation. The coefficients in the mathematical relation are not constant but depend on the so-called time shift operator.

Grey-box storage models are given in [21] and [22]. In the first reference, the parameter values of a serial grey-box model structure are adjusted with a stochastic approximation technique to simulate the oil temperature in the storage of a solar power plant. In the latter work, a grey-box model and its identification procedure is presented for describing the temperature of hot water tanks. The identification procedure is aimed at being a suitable tool for model-based controllers.

Because of the above problems on arduousness and time consumption, this paper intends to establish a simple, general and still precise black-box model, which can be used fast and easily for a wide range of storages. This model is built on a known method of statistics, namely, the multiple linear regression (MLR). After studying the literature, it can be stated that MLR is a missing black-box modelling method in case of storage tanks in spite of the simple applicability. MLR-based models have been worked out in recent works for other working components of hydraulic heating systems, namely, for solar collectors [23] and for pipes [24]. As proposing a new MLR-based storage model, called *LR model* in short, the present one can be considered as the continuation of these works, in the Conclusion of which it was suggested to work out MLR-based models for further elements of heating systems. The contribution of the present paper is the following in details.

1. The linear storage model from Buzás and Farkas [3] that has been already applied successfully in many works in the literature, is slightly modified (according to the short circuit effect) and validated with the measurements on a real solar storage treating it as a separate working component. The gained results support the practical usability of the modified model (as a simple ODE) called *physically-based model* in short below.

2. A general, new and easy-to-apply MLR-based black-box model with low computational demand, called LR model, is proposed for solar storages. The model stands for empirical relations directly between the input variables and the storage temperature and may be the simplest possible black-box model with a rather good precision.

3. Based on measurements, the LR model is validated then compared with the physically-based model with respect to precision. The results show that the LR model is significantly more precise (with a rather low modelling error) than the physically-based model.

It is worth mentioning in advance that the proposed LR model is more easy-to-use and has a much lower computational demand compared to other black-box models. Furthermore, more complex models (e.g. ANNs) can have the same or even lower precision (see e.g. [17]).
The organization of the paper is as follows: Section 2 serves with general features on the studied solar storage type. In Section 3, the physically-based model used for comparison is introduced. The new LR model is worked out in Section 4. Based on measurements and simulated data, Section 5 details the identification and validation of the models, which are compared quantitatively in Section 6. Conclusions and future research proposals close the work in Section 7.

2. General features on the studied solar storages

Fig. 1 shows the schematic view of the studied solar storage type. The storage can be heated up through a heating loop, within which the fluid enters the storage with temperature T_{in} and leaves it with T_{out}. A pump circulates the fluid in the heating loop with 0 or a prefixed constant flow rate value v according to the on/off state of the pump. This working is in accordance with the well-known on/off control, which is likely the most widely used control method in the thermal engineering practice to date (see e.g. [25]) because of its simplicity and that this control method is the nearly optimal or even optimal one in many cases (see e.g. [26]). Sometimes a consumer discharges some fluid from the storage tank with the flow rate v_{load}. The tank, with (geometrical) average inner temperature T_s, has some heat loss to the environment through its insulation.

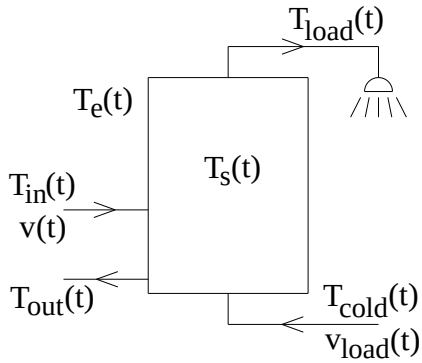

Fig. 1. Schematic view of the solar storages

All time dependent variables T_{cold}, T_e, T_{in}, T_{load}, T_{out}, T_s, v and v_{load} are measured periodically according to a time period Δt.

Below, T_s denotes the geometrical average storage temperature generally, while $T_{s,\mathrm{mod}}$ and $T_{s,meas}$ denote the modelled and measured geometrical average storage temperature, respectively.

3. Physically-based model

Eq. (1) represents the physically-based tank model used in this paper, which is the slightly modified version (completed with a modifying coefficient c_v) of the linear ODE of Buzás and Farkas [3].

$$\frac{dT_{s,\mathrm{mod}}(t)}{dt} = \frac{c_v v(t)}{V}\left(T_{in}(t) - T_{out}(t)\right) + \frac{v_{load}(t)}{V}\left(T_{cold}(t) - T_{load}(t)\right) + \frac{Ak}{\rho c V}\left(T_e(t) - T_{s,\mathrm{mod}}(t)\right) \qquad (1)$$

In Eq. (1), the short circuit effect (mentioned in the Introduction) is considered with the modifying coefficient c_v (which is between 0 and 1) for better modelling precision.

4. LR model

Because of the bounded propagation speed of physical effects and the bounded speed of measurements, the inputs of the LR model are $T_{in}(t-\tau)$, $v_{load}(t-\tau)$ and $T_s(t-\tau)$ with respect to the output (which is the modelled value of $T_s(t)$, that is, $T_{s,\mathrm{mod}}(t)$) at the current time *t*. Here τ is the time delay with respect to the effects of the inlets (more particularly T_{in} and v_{load}) to the interior of the storage tank. Clearly, the previously detected value of the storage temperature (as some initial value of the model) has also essential effect on its current value $T_s(t)$. For simplicity, $T_s(t-\tau)$ is taken as this previous temperature to be considered in the model.

Regarding the solar storage as a black-box, it can be realized that distinct sub-models should be identified as separate parts of the LR model (as a black-box model) for considerably different working conditions. More particularly, the storage tank behaves in an absolutely different way if the pump is on (v>0) or off (v=0) permanently. Namely, under the same initial tank temperature, T_s basically increases if the pump is on and decreases if the pump is off. Furthermore, the effect of T_{in} can be omitted if the pump is permanently switched off as there is no fluid flow into the tank from the heating loop then. Regarding a typical day, when the temperature increase of T_s is high enough (and the consumption load is not extremely high), three different working cases are worth distinguishing in accordance with Fig. 2.

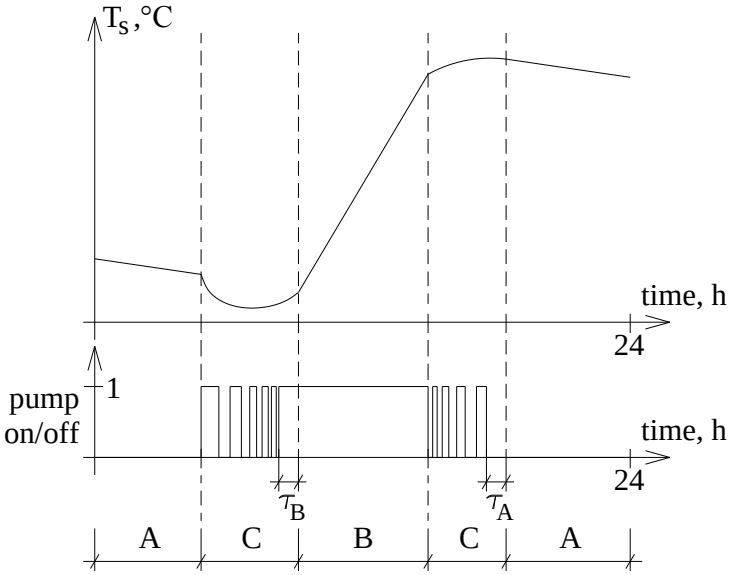

Fig. 2. Solar storage temperature and state of the pump on an average day

Cases A and *B* relate to permanently switched off and switched on pump, respectively, while *Case C* relates to frequent switch-offs and -ons. The specification of all *Cases* can be found below in details (see also Fig. 2).

Case A: Time of permanently switched off pump. More precisely, this *Case* contains the time period, which starts at the beginning of the day and finishes at the first pump switch-on, and each such time period, which starts at a time point when the pump has been continuously switched off for exactly τ_A time and finishes at either the time of the next switch-on or at the end of the day.

Case B: Time of permanently switched on pump. More precisely, this *Case* contains each time period, which starts at a time point when the pump has been continuously switched on for exactly τ_B time and finishes at the next switch-off.

Case C: This *Case* contains all time periods beyond *Cases A* and *B*. (That is, if the current time point does not fit into the above conditions of *Cases A* and *B*, then it belongs to *Case C*.)

Remark

1. Since the fluid in the pipes of the heating loop may be colder than the storage tank at the first switched on working terms of the pump, the heating loop may cool down the tank in the corresponding time period (see the graph of T_s in Fig. 2 just before *Case B*).

2. τ_A is the time which is to be passed after a switch-off to proceed with *Case A*. More particularly, τ_A is the time which is usually enough for T_s to become permanently decreasing (see also Fig. 2), and, this behaviour is attributed to *Case A* as its characteristic feature. Similarly, τ_B is the time which is to be passed after a switch-on to proceed with *Case B*. More particularly, τ_B is the time which is usually enough for T_s to become permanently increasing (see also Fig. 2), and, this behaviour is attributed to *Case B* as its characteristic feature. In fact, it is subjective to some extent to determine the time point when the storage temperature just begins its permanently increasing section (the end of the first section *C* in Fig. 2) and the time point when the temperature just begins its permanently decreasing section (the end of the second section *C* in Fig. 2), that is, to determine the value of τ_A and τ_B. Nevertheless, we can estimate these values based on the (already given) measured temperature graph of more days in the identification (in Section 5.1.2) then the estimated daily values can be averaged, which improves the precision or, alternatively, reduces the roughness of the estimate. Furthermore, rather different estimated values can serve with similarly good results, because of, for example, the following effect. If τ_B is overestimated, the length and significance of *Case C* clearly increases, nevertheless, the number of the time points within *Case C* also increases, which will allow us in the identification process to identify the model parameters with respect to *Case C* more precisely. (Similar considerations hold for τ_A.)

For better modelling precision, separate sub-models are established for each working case based on MLR. In the sub-models for *Cases B* and *C*, T_{in} is considered an input, since some fluid flows into the storage from the heating loop according to the permanent or intermittent pump working. Regarding the sub-model of *Case A*, T_{in} is omitted according to the permanently switched off pump. The schemes of the MLR-based sub-models of each *Case* are presented in Figs. 3 and 4.

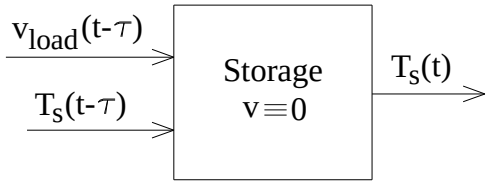

Fig. 3. Scheme of the MLR-based sub-model for *Case A*

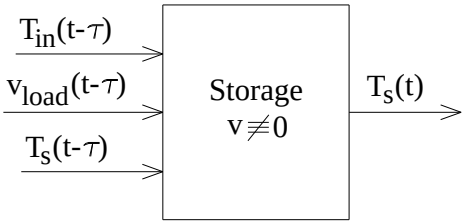

Fig. 4. Scheme of the MLR-based sub-models for *Cases B* and *C*

The LR model is formed with Eqs. (2A), (2B) and (2C), which are simple linear algebraic relations representing the corresponding sub-models of the separate working cases.

Case A: $$T_{s,\mathrm{mod}}(t) = c_{load,A} v_{load}(t-\tau) + c_{s,A} T_s(t-\tau) \quad (2A)$$

Case B: $$T_{s,\mathrm{mod}}(t) = c_{in,B} T_{in}(t-\tau) + c_{load,B} v_{load}(t-\tau) + c_{s,B} T_s(t-\tau) \quad (2B)$$

Case C: $$T_{s,\mathrm{mod}}(t) = c_{in,C} T_{in}(t-\tau) + c_{load,C} v_{load}(t-\tau) + c_{s,C} T_s(t-\tau) \quad (2C)$$

$c_{load,A}$, $c_{s,A}$, $c_{in,B}$, $c_{load,B}$, $c_{s,B}$, $c_{in,C}$, $c_{load,C}$ and $c_{s,C}$ are constant coefficients to be identified.

5. Identification and validation

In this section, the physically-based (Eq. (1)) and the LR (Eqs. (2A), (2B) and (2C)) model are applied to a real storage tank in order to identify and validate them. Then the efficiencies of the models are compared by means of measured and modelled data. For the calculations, the Matlab software [27] has been applied.

In the physically-based model, $T_{s,meas}(0)$ is used as initial value and the measured T_{cold}, T_e, T_{in}, T_{load}, T_{out}, v and v_{load} values are also used for solving Eq. (1) numerically to gain $T_{s,\mathrm{mod}}(t)$. In the identification of the LR model, when the measured values of T_s can be applied, $T_{s,meas}(t-\tau)$ is used as $T_s(t-\tau)$. During the validation of the already identified LR model, the previously modelled value $T_{s,\mathrm{mod}}(t-\tau)$ is used as $T_s(t-\tau)$ when modelling $T_s(t)$. Measured $T_{in}(t-\tau)$ and $v_{load}(t-\tau)$ values are available both during the identification and the validation. According to the specification of Δt, the measurements happen at times $t = 0, \Delta t, 2\Delta t, 3\Delta t, \ldots$ Practically, the modelled value of T_s (that is $T_{s,\mathrm{mod}}$) is determined in the LR model also at times $t = 0, \Delta t, 2\Delta t, 3\Delta t, \ldots$ Furthermore, for simplicity, $\tau = \Delta t$ is assumed in the LR model. *Case A* holds and $T_s(t-\tau) = T_{s,meas}(0)$ is used as measured initial condition in Eq. (2A) at $t = \tau$ (at the beginning of the day).

The modelled real storage tank is the solar storage of a measured solar heating system [28] installed at the Szent István University (SZIU) in Gödöllő, Hungary. This storage tank will be called *SZIU storage* in short. Fig. 5 shows the photo of the SZIU storage, which contains preheated domestic water for a kindergarten at the campus of the university.

Fig. 5. SZIU storage [29]

The heat is transferred from a solar collector field into the SZIU storage by means of a heating loop equipped with a pump working in on/off working. Since the heating loop contains rather long (more than 100 metres long) pipes, the cooling down effect mentioned in the *Remark* may be significant. The measurements are carried out once a minute, that is, Δt =1 min. The volume of the SZIU storage is 2 m^3. Based on the observation of the measured storage temperature of the whole four days for the identification (see Section 5.1), τ_A and τ_B are estimated 10 min (see also Note 2 in the *Remark*). These and other important parameter values can be found in Table 1.

Table 1. Parameter values of the SZIU storage and the models

	Physically-based model	LR model
A, m^2	4	-
c, $Jkg^{-1}K^{-1}$	4200	-
c_v, - (identified)	0.60	-
k, $Wm^{-2}K^{-1}$ (identified)	2.87	-
V, m^3	2	-
Δt, s	60	60
ρ, kgm^{-3}	1000	-
τ, s	-	60
τ_A, s	-	600
τ_B, s	-	600
$c_{load,A}$, Ksm^{-3} (identified)	-	3.6290
$c_{s,A}$, - (identified)	-	0.9998
$c_{in,B}$, - (identified)	-	0.0044
$c_{load,B}$, Ksm^{-3} (identified)	-	11.2829
$c_{s,B}$, - (identified)	-	0.9958
$c_{in,C}$, - (identified)	-	0.0007

$c_{load,C}$, Ksm^{-3} (identified)	-	24.6179
$c_{s,C}$, - (identified)	-	0.9994

There is a temperature sensor inside the upper and another one in the lower third of the storage, the values of which are averaged geometrically and considered as the measured value of T_s, that is, $T_{s,meas}$. The pump flow rate v is 0 or 0.55 m^3/h (based on measured data).

The following indices (corresponding to the currently investigated day) are used in this paper for the evaluation. The *average of error* is the time average of $\left(T_{s,\mathrm{mod}} - T_{s,meas}\right)$, the *average of absolute error* is the time average of the absolute value $\left|T_{s,\mathrm{mod}} - T_{s,meas}\right|$. The *average of absolute error in %* is the average of absolute error divided by the (positive) difference between the (daily) maximal and minimal value of $T_{s,meas}$.

5.1. Identification of the models

For the identification, the measured data of 4 days are selected in such a way that they represent a wide range of possible working conditions in a selected season (in summer). Two of the days (8th June, 2012; 28th June, 2012) are with relatively high consumption load (more than 1000 litres) and two ones (24th June, 2012; 2nd July, 2012) are with relatively low consumption load (less than 200 litres). Based on numerous computer experiments (cannot be detailed here), such 4 days has proved to be satisfactory for the identified model to have a rather good accuracy. Practically, the 4 days are selected from the first third of the summer. In this way, the already identified model can be conveniently used in the remaining two summer months. (To apply the model for a year, the identification could be carried out for all seasons separately to achieve maximal yearly precision.)

T_{cold}, T_e, T_{in}, T_{load}, T_{out}, v and v_{load} are assumed to be piecewise constant in the physically-based model in accordance with the time step of the measurements.

5.1.1. Identification (physically-based model)

The parameters k and c_v are identified in the physically-based model. First, k is identified based on such a (measured) time period when both the pump flow rate and the consumption load are 0. In this case, it is assumed that the heat loss to the environment is the only phenomenon which is responsible for the temperature change (decrease) of the storage, so the value of k can be directly set by fitting the rate of the measured temperature decrease to the rate of the modelled one. This process has been carried out based on a section of the final part of the day with decreasing T_s on 2nd July, 2012. More precisely, this time period is between about 18 and 21 hours on the mentioned day. See the left-hand side of Fig. 7, where it can be roughly seen that the measured and modelled curves of T_s are really parallel in the corresponding period. Then the value of c_v is set in such a way that the mean % value of the average of absolute error is minimal with respect to the 4 identified days. The such identified k and c_v values can be found in Table 1.

Table 2 presents the average of error and the average of absolute error values for two days (28th June, 2012; 2nd July, 2012) of the identification (with the already identified physically-based model). The average of absolute error in % is also given for both days. The mean of the % values relating to all 4 days of the identification can be also seen in Table 2 (11.6 %).

Table 2. Average of error and average of absolute error for the models

	Physically-based model	LR model

Identification	28th June	Average of error	-0.46 °C	0.07 °C
		Average of absolute error	0.58 °C; 10.2%	0.15 °C; 2.7%
	2nd July	Average of error	-0.17 °C	-0.03 °C
		Average of absolute error	0.34 °C; 3.6%	0.13 °C; 1.4%
	Mean % value for the whole identification (4 days)	Average of absolute error	11.6%	2.8%
Validation	3rd August	Average of error	0.25 °C	-0.02 °C
		Average of absolute error	0.26 °C; 3.5%	0.11 °C; 1.4%
	5th August	Average of error	0.51 °C	-0.20 °C
		Average of absolute error	0.59 °C; 13.9%	0.22 °C; 5.3%
	Mean % value for the whole validation (3rd July – 31st August)	Average of absolute error	16.1%	6.2%

On the left-hand side of Figs. 6 and 7, the modelled and measured tank temperatures are compared for two days (28th June, 2012; 2nd July, 2012) of the identification of the physically-based model. The pump state (switched on or off) is also presented in both figures.

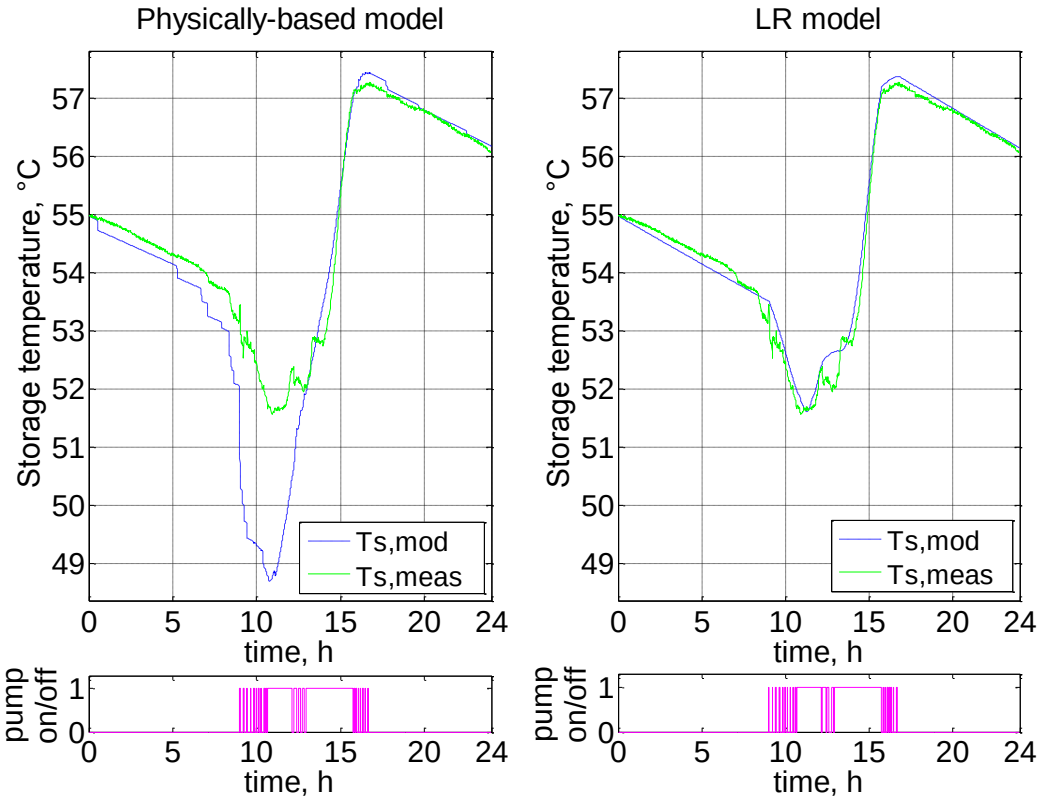

Fig. 6. Modelled $T_{s,\mathrm{mod}}$ and measured $T_{s,meas}$ storage temperatures on 28th June, 2012 for the physically-based and LR models (identification)

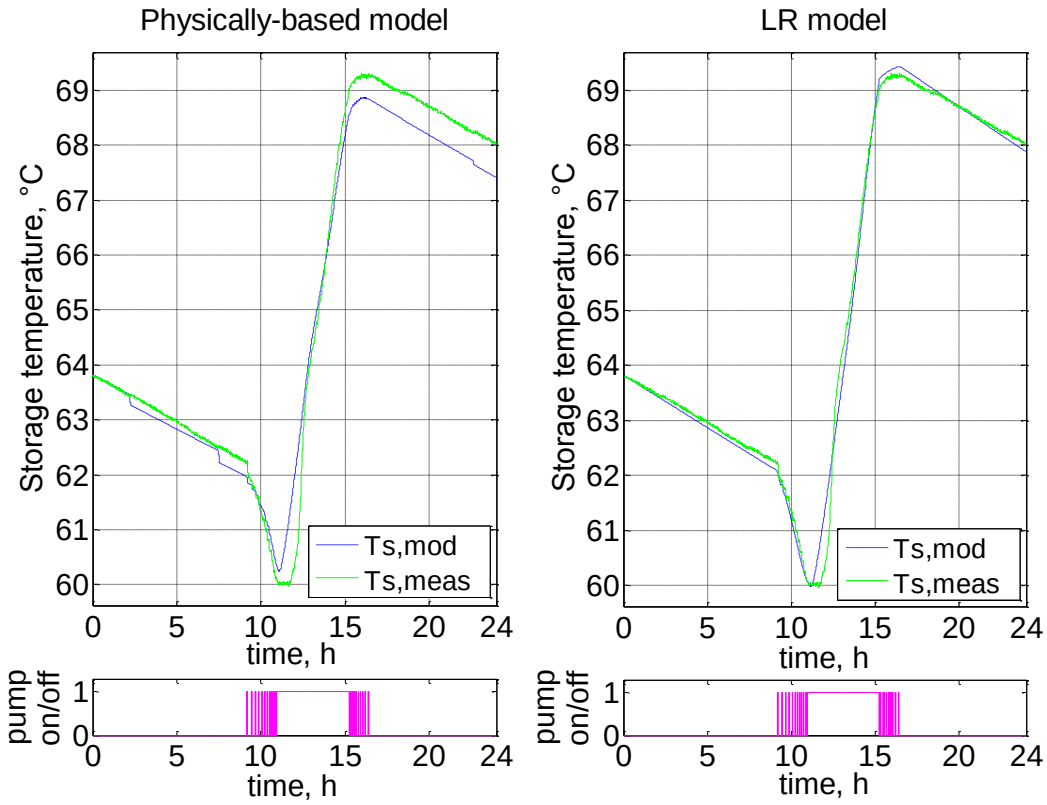

Fig. 7. Modelled $T_{s,\mathrm{mod}}$ and measured $T_{s,meas}$ storage temperatures on 2nd July, 2012 for the physically-based and LR models (identification)

5.1.2. Identification (LR model)

Three standard independent MLR routines are used based on the measured data of all distinct working cases of the LR model (*Cases A, B,* and *C*) to identify parameters $c_{load,A}$, $c_{s,A}$, $c_{in,B}$, $c_{load,B}$, $c_{s,B}$, $c_{in,C}$, $c_{load,C}$ and $c_{s,C}$ in Eqs. (2A), (2B) and (2C) in Section 4. For the identification, the measured data of all *Cases* are selected from the same 4 identified days as in case of the physically-based model in Section 5.1.1. The known standard MLR routine (on the basis of least squares minimization) is built in most statistical or spreadsheet programs (SPSS, Excel, etc.), so it is not needed to be detailed. The identified parameters of the LR model are given in Table 1. The results of an MLR routine is generally evaluated with the square of the correlation coefficient R^2. Table 3 presents the R^2 values for all working cases corresponding to the whole identification.

Table 3. R^2 values for the LR model

Mode	A	B	C
R^2	0.99999	0.99998	0.99997

R^2 is close to 1 in all of *Cases A, B* and *C,* because of which the LR model can be considered very reasonable. The indices of Table 2 are more expressive and more important, especially, in the aspect of the comparison with the physically-based model applied already successfully in the literature. Table 2 presents the average of error and the average of absolute error values for two days (28th June, 2012; 2nd July, 2012) of the identification (with the already identified LR model). The average of absolute error in % is also given for both days. The mean of the % values relating to all 4 days of the identification can be also seen in Table 2 (2.8 %).

On the right-hand side of Figs. 6 and 7, the modelled and measured tank temperatures are compared for two days (28th June, 2012; 2nd July, 2012) of the identification of the LR model. The working state of the pump is also presented in the figures.

5.2. Validation of the models

For validation, the identified physically-based model and the identified LR model are used with the proper measured inputs from the remaining two months of the summer. More precisely, one of the inputs of the LR model is changed compared to the inputs of the identification, that is, the modelled $T_{s,\mathrm{mod}}(t-\tau)$ is applied as $T_s(t-\tau)$ in the LR model (2A), (2B) and (2C) (not $T_{s,meas}(t-\tau)$). The modelled period for the validation, from 3rd July, 2012 to 31st August, 2012, contains 56 days (according to some short technical interruptions).

The modelled and the measured storage temperatures are compared and evaluated in case of the two models. Table 2 presents the average of error and the average of absolute error values for two days (3rd August, 2012; 5th August, 2012) of the validation for the models. The average of absolute error in % is also given for both days. The mean of the % values relating to the whole time period of 3rd July – 31st August can be also seen in Table 2 (16.1% and 6.2% for the physically-based and LR models, respectively).

In Figs. 8 and 9, the modelled and measured tank temperatures are compared in case of both models for two days (3rd August, 2012; 5th August, 2012) of the validation. The pump state is also presented in both figures.

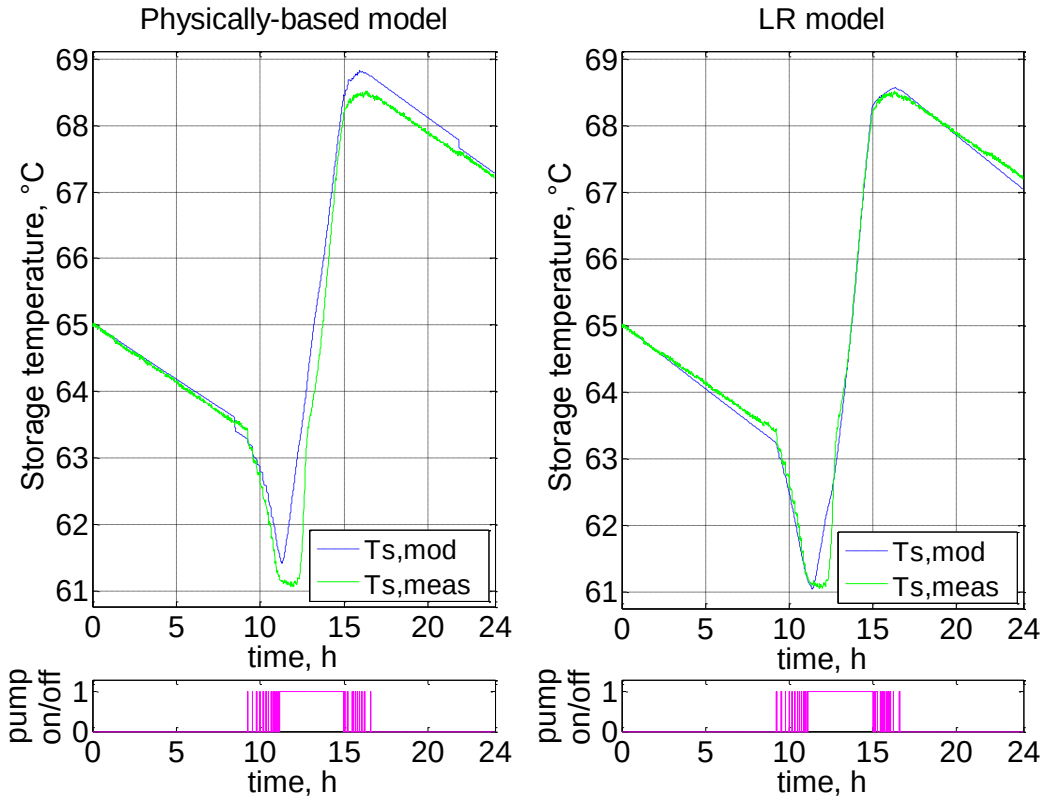

Fig. 8. Modelled $T_{s,\mathrm{mod}}$ and measured $T_{s,meas}$ storage temperatures on 3rd August, 2012 for the physically-based and LR models (validation)

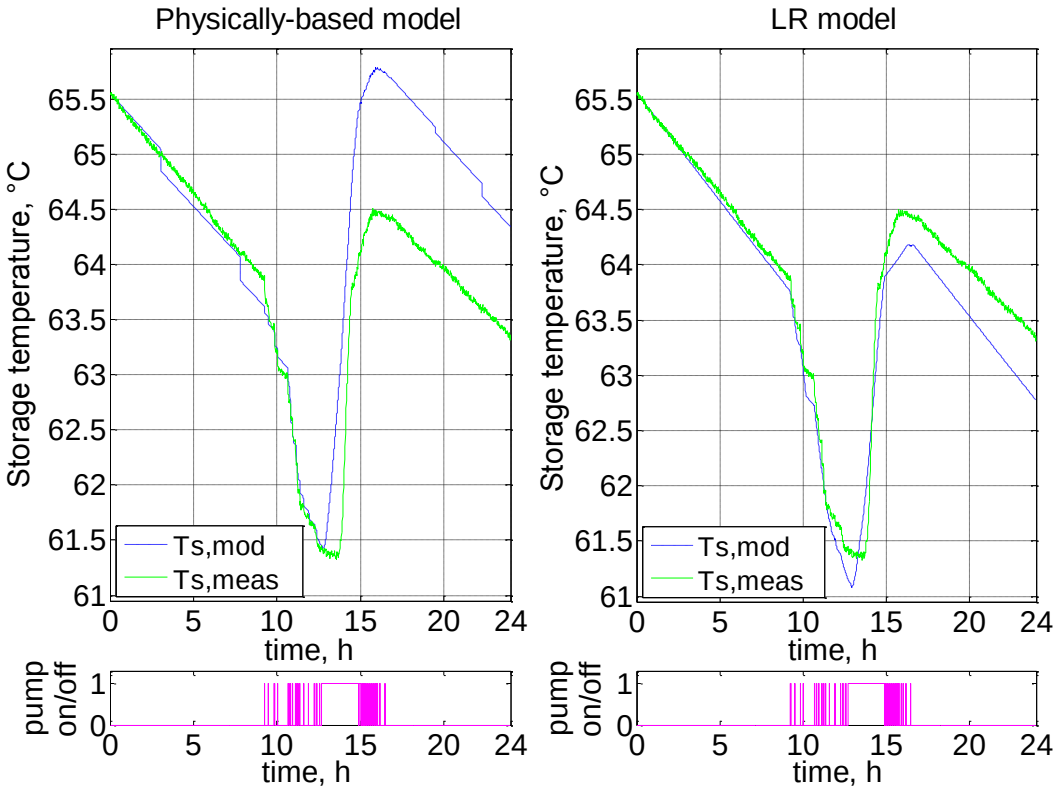

Fig. 9. Modelled $T_{s,\mathrm{mod}}$ and measured $T_{s,meas}$ storage temperatures on 5th August, 2012 for the physically-based and LR models (validation)

6. Comparison and discussion

The LR model can predict the storage temperature rather precisely, particularly, more precisely than the physically-based model. According to the validation, the mean (absolute) modelling error (in %) is 6.2% with the LR model and 16.1% with the physically-based model. The latter has been already used successfully in many works in the literature. Thus the accuracy of the LR model can be stated very well regarding general solar thermal engineering purposes (studying and developing the thermal processes of solar storages).

In addition, the R^2 values between the modelled and measured data with respect to the above selected 4 days (2 identified and 2 validated ones) have been determined for the models. They are given in Table 4 confirming the better precision of the LR model compared to the physically-based one.

Table 4. R^2 values for the models for whole days

Date	28th June, 2012 (identification)	2nd July, 2012 (identification)	3rd August, 2012 (validation)	5th August, 2012 (validation)
Physically-based model	0.918	0.990	0.976	0.713
LR model	0.988	0.995	0.994	0.965

It should be mentioned that the storage temperature cannot be expected to be predicted perfectly with relatively simple models, largely because of the above discussed short circuit effect. This effect must be rather complex, and thus hard to model, depending on the current flow rate values v, v_{load} and, probably, on the different temperature values of the system. In case of the SZIU storage, v is relatively high (0.55 m^3/h) compared to the tank volume (2 m^3), which likely makes the short circuit effect significant. It should be also mentioned that

neither model takes into account the thermal stratification inside the storage, which must be more disadvantageous for the physically-based model than for the LR model, since the strength of black-box models is just that they can often serve with rather precise empirical relations without dealing with the physical backgrounds. These must be the reason for that the accuracy of the physically-based model is not outstanding (the error is over 10%). On the other hand, these difficulties also confirm the reasonability and robustness of the LR model considering that its error is rather low (close to 5% in the validation). In sum, the accuracy of the LR model can be stated very well, especially, regarding its simplicity. Furthermore, the physically-based model can be stated still usable in the practice as a simple ODE model (c.f., for example, the 30% and 10% error values reported in [13] in the Introduction).

In fact, there are more parameters to be identified in the LR model than in the physically-based model, but, on the whole (and even in the identification), it is much easier and faster to apply the LR model, since it means only one explicit algebraic operation at each minute (each time step) as it is a simple algebraic equation (depending on the current *Case*), while the physically-based model is a differential equation, which must be discretized then solved numerically. Generally, at least 100 time steps is suggested per minute for a satisfactorily precise numerical calculation, which means at least 100 times more algebraic operations than in case of the LR model.

7. Conclusions

The purpose of the present research was to work out a general, easy-to-apply but still precise mathematical model for solar storages. After studying the literature, it can be stated that MLR is a missing black-box modelling method in case of hydraulic storage tanks in spite of the (linear algebraic) simplicity. In this paper, a new model, called LR model, has been established and validated based on measured data to fulfil this research gap. The proposed LR model may be the simplest possible black-box model with a rather low modelling error (close to 5%, more precisely, 6.2%) and with very low computational demand.

The slightly modified version of the physically-based linear ODE model of Buzás and Farkas [3] for solar storages has been validated as well with a modelling error of 16.1%. This model accounts for the short circuit effect in storage tanks.

Considering the complex, and hard to predict, short circuit effect of the modelled solar storage (the SZIU storage), the accuracy and the robustness of the LR model can be stated very well, furthermore, the physically-based model is still usable in the practice as a simple ODE model. It is not difficult to think that both models would be even more accurate in case of a more regular tank (with less significant short circuit effect).

After assigning *Cases A*, *B* and *C*, it is not difficult to identify the LR model (2A), (2B) and (2C) for any storage tanks according to the identification method of Section 5.1.2 (in case of similar inputs and output as there), so the LR model is also general.

Basically, the LR model can be used for very fast but still precise storage modelling/simulation, nevertheless, because of its simple linear algebraic equations, the computational demand is very low, which may make it convenient for model-based control schemes. The advantage of the simple usability and the low computational demand can be seen especially compared to other, more complex black-box type models having essentially the same or even lower precision.

MLR-based models have been worked out in recent works for other working components of hydraulic heating systems, namely, for solar collectors [23] and for pipes [24]. The LR model can be considered as the continuation of these works, in the Conclusion of which it was suggested to work out MLR-based models for further elements (like storage tanks) of hydraulic heating systems.

Future researches may deal with connecting the already worked out MLR-based models of the separate working components to form an easy-to-use and precise MLR-based model for

complete solar heating systems. Furthermore, new model-based controls could be created on the basis of the LR model.

Acknowledgement

The author thanks the Editor for the help in submitting the paper, the Department of Physics and Process Control in the Faculty of Mechanical Engineering (SZIU) for the measured data of the SZIU storage and his colleagues, especially Éva Dékány, in the Department of Mathematics for the support.

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

References

[1] Duffie JA, Beckman WA (2006) Solar engineering of thermal processes, 3rd edn. John Wiley and Sons, New York

[2] Zeghib I, Chaker A (2011) Simulation of a solar domestic water heating system. Energy Procedia 6:292-301.

[3] Buzás J, Farkas I (2000) Solar domestic hot water system simulation using blockoriented software. In: The 3rd ISES-Europe solar world congress (Eurosun 2000), Copenhagen, Denmark, CD-ROM Proceedings, p. 9.

[4] Kumar R, Rosen MA (2010) Thermal performance of integrated collector storage solar water heater with corrugated absorber surface. Appl Therm Eng 30:1764-1768.

[5] Jannatabadi M, Taherian H (2012) An experimental study of influence of hot water consumption rate on the thermal stratification inside a horizontal mantle storage tank. Heat Mass Transf 48(7):1103-1112.

[6] Duer M (2011) Passive mixing systems improve storage tank water quality. Opflow 37(8):20-23.

[7] Han YM, Wang RZ, Dai YJ (2009) Thermal stratification within the water tank. Renew Sustainable Energy Rev 13:1014-1026.

[8] Bayón R, Rojas E (2013) Simulation of thermocline storage for solar thermal power plants: From dimensionless results to prototypes and real-size tanks. Int J Heat Mass Transfer 60:713-721.

[9] Altuntop N, Arslan M, Ozceyhan V, Kanoglu M (2005) Effect of obstacles on thermal stratification in hot water storage tanks. Appl Therm Eng 25:2285-2298.

[10] Zachár A, Aszódi A (2007) Numerical analysis of flow distributors to improve temperature stratification in storage tanks. Numer Heat Tr A-Appl 51:919-940.

[11] Zachár A (2013) Investigation of a new tube-in-tube helical flow distributor design to improve temperature stratification inside hot water storage tanks operated with coiled-tube heat exchangers. Int J Heat Mass Transfer 63:150–161.

[12] Wang Z, Zhang H, Dou B, Huang H, Wu W, Wang Z (2017) Experimental and numerical research of thermal stratification with a novel inlet in a dynamic hot water storage tank. Renew Energy 111:353-371.

[13] Johannes K, Fraisse G, Achard G, Rusaouën G (2005) Comparison of solar water tank storage modelling solutions. Sol Energy 79:216-218.

[14] Klein SA et al. (2005) TRNSYS 16 – A transient system simulation program. Solar Energy Laboratory, University of Wisconsin-Madison

[15] Meyer JP, Raubenheimer PJA, Kruger E (2000) The influence of return loop flow rate on stratification in a vertical hot water storage tank connected to a heat pump water heater. Heat Transfer Eng 21(2):67-73.

[16] Kalogirou SA, Panteliou S, Dentsoras A (1999) Modelling of solar domestic water heating systems using artificial neural networks. Sol Energy 65(6):335-342.

[17] Kalogirou SA (2000) Applications of artificial neural-networks for energy systems. Appl Energy 67:17-35.

[18] Géczy-Víg P, Farkas I (2010) Neural network modelling of thermal stratification in a solar DHW storage. Sol Energy 84:801-806.

[19] Cabello JM, Cejudo JM, Luque M, Ruiz F, Deb K, Tewari R (2011) Optimization of the size of a solar thermal electricity plant by means of genetic algorithms. Renew Energy 36(11):3146-3153.

[20] Romero JA, Navarro-Esbrí J, Belman-Flores JM (2011) A simplified black-box model oriented to chilled water temperature control in a variable speed vapour compression system. Appl Therm Eng 31:329-335.
[21] Arahal MR, Cirre CM, Berenguel M (2008) Serial grey-box model of a stratified thermal tank for hierarchical control of solar plant. Sol Energy 82:441-451.
[22] De Ridder F, Coomans M (2014) Grey-box model and identification procedure for domestic thermal storage vessels. Appl Therm Eng 67(1-2):147-158.
[23] Kicsiny R (2016) Improved multiple linear regression based models for solar collectors. Renew Energy 91:224–232.
[24] Kicsiny R (2017) Grey-box model for pipe temperature based on linear regression. Int J Heat Mass Transfer 107:13–20.
[25] Araújo A, Pereira V (2017) Solar thermal modeling for rapid estimation of auxiliary energy requirements in domestic hot water production: On-off flow rate control. Energy 119:637-651.
[26] Winn CB, Hull DE (1979) Optimal controllers of the second kind. Sol Energy 23:529-534.
[27] Etter DM, Kuncicky D, Moore H (2004) Introduction to MATLAB 7. Springer
[28] 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 Frankovic B/, Croatian Solar Energy Association, Opatija, 2000., 81-88.
[29] Géczyné Víg P (2007) Modelling of solar heating systems with neural network. Dissertation, Szent István University, Gödöllő, p. 130. [in Hungarian] https://szie.hu//file/tti/archivum/geczine_v_p_phd.pdf Accessed 10 August 2017