Heuristic algorithm for the piecewise linear segmentation of multiple time-series for solar thermal systems inverse modelling

Vitor V. Lopes[*,1], Filipa Ferro[1], Maria J. Carvalho[2] and Augusto Q. Novais[1]
[1]Energy Systems Modeling and Optimization Unit, LNEG, 1649-038 Lisboa, Portugal
[2]Solar Energy Lab, LNEG, 1649-038 Lisboa, Portugal

Abstract

This paper presents a novel algorithm for the piecewise linear segmentation of multivariate time-series and proposes its application to the analysis of hot water thermal solar systems (TSS). The ISO 9459-5:2007 norm describes a non-intrusive dynamic test for the performance assessment of TSS. This allows to characterize the system heat losses and the thermal stratification properties, as well as to predict its long-term performance. The application of this norm requires an inverse modeling approach where the parameters of a simplified plug flow storage model, based on simulation runs, are determined through an optimization procedure aiming at the adjustment of the predicted results to those obtained by a predefined experimental test sequence (3-5 days). This paper proposes a new method to decrease the computation time required for the model simulation, which is based on the segmentation of the multivariate time-series into a piecewise linear approximation, where the number of segments is critically selected. An illustrative example is presented consisting in the simulation of a real 3-day experimental dataset with 26873 points and a 15 s sampling rate.

Keywords: piecewise linear approximation, thermal solar systems, partial differential equation, inverse modeling

1. Introduction

The use of thermal solar systems (TSS) in buildings to supply domestic hot water and/or space heating needs an energy storage device to deal with the imbalance between the requested demand and available supply. The TSS performance is hence related to its capability to handle disturbances in the incoming solar radiation energy and the short-term intakes of cold water, while maintaining the output temperature. Performance improves with a correctly dimensioned thermal energy storage (TES), designed to develop a good thermal stratification and be well insulated. Temperature stratification occurs due to density differences between hot and cold water and can be destroyed by thermal conduction along storage tank walls, processes that induce turbulence such as plume entrainment, natural convection and inlet jet mixing.

The TSS performance assessment is based on a non-intrusive dynamic system test defined by the ISO 9459-5:2007 norm [1], which aims at the prediction of the long term

* Author to whom all correspondence should be addressed. E-mail: vitor.lopes@lneg.pt

behavior of a given TSS using a simplified 1D nonlinear partial differential equation (PDE) model [2]. The model parameters are estimated from the experimental data collected during a predefined test procedure applied to the TES under analysis. This test is performed outdoors during 3 to 5 days and is designed to mimic a typical utilization pattern. The high sampling frequency specified by the ISO norm results in large datasets and hence long CPU times, even with the use of efficient integration software codes [3].
This paper addresses this issue by proposing a heuristic segmentation algorithm to derive a piecewise linear approximation (PLA) of the dataset, based on a new algorithm for the automatic definition of an irregular time grid in multivariate data-sets employing a multiple objective function. The PLA was favored since it is widely used among the methods to solve PDE representing spatial domain solutions [4].
The paper is structured in three main sections: next section presents the simplified nonlinear PDE model for the TSS and briefly explains the PDE solution strategy. The heuristic segmentation algorithm follows and the methodology for the selection of the number of integration points is explained. Finally, an illustrative case is presented of the impact of the number of integration points upon the simulation results.

2. Modeling of the thermal solar system

The behavior of thermal solar systems may be approximated by the nonlinear PDE defined by Eq.1. The model developed by Spirkl et al.[2] supports the TSS performance characterization defined by the ISO 9459-5:2007 norm [1]. It describes the temperature profile $T(t,h)$ as a function of the normalized storage height (h) and time (t):

$$C_S\frac{\partial T(t,h)}{\partial t}=\underbrace{U_S(T-T_{SA})}_{term\,1}\underbrace{-\dot{C}_S\frac{\partial T}{\partial h}+\frac{\partial}{\partial h}\left[D_L\dot{C}_S\frac{\partial T}{\partial h}\right]}_{term\,2}+\underbrace{\frac{\partial}{\partial h}\left[b\exp\left(-z\frac{\partial T}{\partial h}\right)\frac{\partial T}{\partial h}\right]}_{term\,3}+Q(t) \quad (1)$$

where C_S is the TES thermal capacity (JK^{-1}), U_S is the overall heat loss coefficient for the TES (WK^{-1}), T_{SA} is the ambient temperature measured at the TES location (K), $\dot{C}_S$ is the intake water energy capacitance rate (WK^{-1}) which equals the product of the mass flow and water thermal capacity, D_L is the draw-off mixing coefficient, b and z are two parameters that describe the thermal diffusion and stratification process, respectively. The model accounts for the TES heat losses (term.1); a plug-flow model component, which describes the water draw-off impact upon the temperature profile (term.2); a diffusion term dependent on the temperature gradient to model the thermal stratification (term.3); and also a time-varying energy source/sink term, $Q(t)$ defined by Eq. (2):

$$Q=\underbrace{\delta(h)A_c^{\bullet}\left[G_T^{\bullet}-U_c^{\bullet}(T-T_{CA})\right]^{+}}_{term\,4}+\underbrace{\dot{C}_S\delta(h)(T_{CW}-T)}_{term\,5} \quad (2)$$

where $\delta(x)$ is the unit Dirac delta function, $[x]^{+}$ is the Heaviside operator which equals zero for $x\leq 0$ and x otherwise, $A_c^{\bullet}$ is the effective collector area (m^2), $G_t^{\bullet}$ is the measured solar radiation power (W), $U_c^{\bullet}$ is the effective heat loss coefficient (Wm^{-2}K^{-1}) of the solar collector loop, $T_{out}(t)=T(t,1)$ is the TES water draw-off temperature (K), and T_{CA} and T_{CW} are the ambient temperatures measured at the collector location and water inlet point, respectively. In this equation $Q(t)$ combines two different contributions: the net power provided by the solar collector described by the Hotter-Whillier-Bliss equation [5] (term.4) and the power provided by the water intake (term.5). These two terms are modeled as point sources/sinks located at $h=0$ (bottom of the TES).
It is also assumed that: a) no heat transfer occur in the TES top and bottom interfaces; b) the initial temperature profile in the storage is constant, which is true for the data

collected using ISO norm test protocol. These conditions define in Eq.3 the initial condition and the two boundary conditions, required for the complete specification of the nonlinear PDE problem.

$$\left.\frac{\partial T\,(t,h)}{\partial t}\right|_{h=0} = \left.\frac{\partial T\,(t,h)}{\partial t}\right|_{h=1} = 0, \forall_{t>0} \quad T\,(0,h) = T_0, \forall_{h\in]0,1[} \tag{3}$$

The numerical solution for this nonlinear PDE is implemented using the spectral method. The Galerkin method [4] allows to transform the original PDE into a nonlinear differential algebraic equation (DAE), by writing the solution estimate in the following form:

$$T\,(t,h) = \sum_{i=0}^{N-1} a_i(t)\phi_i(h) \tag{4}$$

where $a_i(t)$ are the temporal dependent coefficients and $\phi_i(h)$ are the set of Chebyshev polynomial basis defined in the spatial domain [6]. The PDE weak form of Eq.1 can be written as follows:

$$\int_0^1 w_i(h)\left(C_S\frac{\partial T(t,h)}{\partial t} + U_S(T-T_{SA}) + \dot{C}_S\frac{\partial T}{\partial h}\right)dh - \int_0^1\frac{\partial w_i(h)}{\partial h}\left(D_L\dot{C}_S\frac{\partial T}{\partial h} + b\exp\left(-z\frac{\partial T}{\partial h}\right)\frac{\partial T}{\partial h}\right)dh \tag{6}$$

$$-D_L\dot{C}_S\left(w_i(h)\frac{\partial T}{\partial x}\right)\Big|_0^1 - b\left(w_i(h)\exp\left(-z\frac{\partial T}{\partial x}\right)\frac{\partial T}{\partial x}\right)\Big|_0^1 - \int_0^1 w_i(h)Q(t)\,dh = 0$$

where the weight functions are defined by $w_i(h) = \phi_i(h)$, and:

$$\int_0^1 w_i(h)Q(t)\,dh = w_i(0)\cdot\left(A_c^*\left[G^*(t) - U_c^*(T(0,t) - T_{CA}(t))\right]^+ + \dot{C}_S(t)(T_{CW}(t) - T(0,t))\right) \tag{7}$$

The nonlinear DAE system is constructed by evaluating the integrals for the different weighting functions. By making use of the Chebyshev polynomials properties [6], it is possible to compute the integrals using closed-form solutions for all terms except the diffusion. In this case, the integral is evaluated numerically by using the Gauss quadrature rule [7].

The TSS performance analysis requires the identification of the seven model parameters (C_S, A_c^*, U_c^*, U_s^*, D_L, b and z) from the data recorded in the experimental test protocol, i.e. $G_t^*(t)$, $\dot{C}_s(t)$, $T_{SA}(t)$, $T_{CA}(t)$, $T_{CW}(t)$ and $T_{out}(t)$. This is achieved by minimizing the least squares deviations between the TES output temperatures and the corresponding experimental values. The nonlinear DAE system solution is computed numerically using the sundials package [3] for a given set of predefined integration points.

Prior to the data fitting step the experimental data, with the exception of water temperatures, require preprocessing to remove high frequency perturbations and invalid data points. The air measured temperatures (T_{SA} and T_{CA}) are filtered with a mean filter of 50 min. The case of the solar power is more complex since it can be affected by both fast or long perturbations caused by the presence of different cloud types: since fast perturbations are considered to have a negligible impact upon the TSS performance, a 15 min median filter is used.

3. An heuristic algorithm for the piecewise linear approximation of multivariate time-series

This section proposes to describe the experimental dataset by a multivariate piecewise linear approximation (PLA). The rationale is that by reducing the number of integration points required to compute the solution, the time required for the TSS simulation will also decrease. The PLA requires the definition of the linear approximation parameters,

as well as the specification of the branch changing points (breakpoints). Although it is possible to express this mathematically as a mixed integer linear programming problem, its resolution is computational intensive [8]. The proposed heuristic algorithm is based on the extension of the bottom-up algorithm used for time-series segmentation [8] to the handling of multivariate data.

In short, the univariate version of the algorithm can be described as follows. Firstly, the initialization is performed with the finest possible linear approximation, formed by the segments between the k^{th} and $(k+1)^{th}$ points. For all segments, the error of joining the two adjacent segments is computed using a linear regression approximation. Next, the two segments with the smaller error are selected. If the error is below a pre-specified threshold then these are grouped. The procedure is then repeated until no more segments can be selected.

The multivariate version of the algorithm uses a multivariate linear regression for the error computation in each segment and redefines the objective function as a weighted sum of the individual variable errors. This transforms the original problem to a multi-objective optimization problem in which the weight selection plays a key role for balancing the quality of the PLA for all the variables in the dataset.

The weights are selected using two different approaches depending on the measured variable type. For the temperature sensors, the corresponding weights are fixed to the measurement errors established by the ISO norm, i.e. 0.5 K and 0.3 K for the air and liquid temperatures, respectively. For the solar radiation (G_t^*) and load capacitance rate ($\dot{C}_s$), the weights are determined using the following procedure. Firstly, apply the univariate version of the segmentation algorithm to each time-series for a different number of approximation errors. Then, draw the curve of the approximation error as a function of the number of segments and find the curve inflection point (curve knee). For TSS data, the weights are typically between 1-1.5% of the maximum value recorded of the corresponding time-series.

4. Results and discussion

For illustration purposes, this section presents a comparison study to analyze the impact of the proposed piecewise linear approximation on the simulation of a real 3-day experimental dataset with 26873 points and a 15 s sampling rate. The study compares two different approximation alternatives: the multivariate PLA (mPLA) and the application of the univariate PLA independently to each time-series (uPLA). For this dataset, the weights for the mPLA algorithm are set to 20W, $10WK^{-1}$, 0.5K, 0.5K, 0.3K, 0.3K for the G_t^*, $\dot{C}_s$, T_{SA}, T_{CA}, T_{CW} and T_{out}, respectively. The error criteria required for the application of the mPLA are selected using the inflection point criteria explained earlier, based on the tradeoff between the approximation error and number of segments.

The analysis of Table.1 shows that it is possible to compress significantly the experimental dataset without incurring in large approximation errors. The results support the effectiveness of the mPLA approach in further decreasing the number of breakpoints. From Table.2 and Fig.1 it is found that the simulation with the mPLA provides essentially the same dynamic behavior as with the full dataset, while the CPU time decreases significantly and also that for uPLA and mPLA similar simulation results are obtained, but with the latter requiring about 50 percent less CPU time.

	uPLA		*mPLA*	
Time Series	Max. Error	N° Pieces	N° Pieces	Max. Error
G_t*	20 W	338	498	19.81 W
$\dot{C}_s$	10 WK^{-1}	132		9.71 WK^{-1}
T_{CA}	0.5 K	56		0.48 K
T_{SA}	0.5 K	23		0.4 K
T_{CW}	0.3 K	122		0.29 K
T_{out}	0.3 K	101		0.28 K

Table.1- Performance of the two PLA alternatives in terms of the number of branches and maximum approximation error for each variable in the dataset

	a)	b)	c)
Points	26873	498	772
CPU time [s]	5011	16	28
Max. Error [K]	-----	0.5	0.05
Sum Abs. Error [K]	-----	0.5	0.05

Table.2- Simulation differences in the top temperature between the original dataset (a), the uPLA (b) and the mPLA (c).

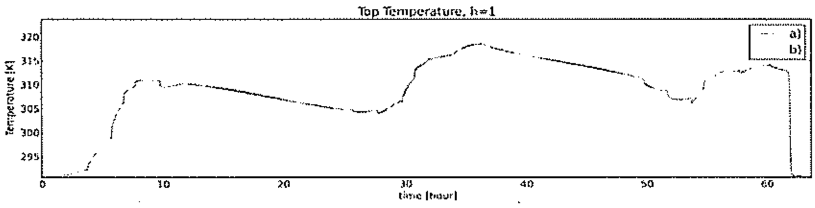

Fig. 1: Temporal evolution of the TES top temperature for the two simulation cases: a) using all the dataset points (26873 integration points); b) using mPLA of the original dataset (498 integration points).

5. Conclusions

The paper proposes the use of a heuristic multivariate piecewise linear approximation algorithm for the compression of experimental datasets used for the performance analysis of thermal solar systems. Results show that it is possible to reduce the number of integration points from 26873 to 498 without affecting the simulation results, while decreasing the simulation time from 5011 down to 16 seconds.

6. Acknowledgment

Work was partially supported by the FCT project PTDC/ENR/70844/2006.

References

[1] ISO 9459-5:2007, Solar heating - Domestic water heating systems - Part 5: System performance characterization by means of whole-system tests and computer simulation, International Organization for Standardization, Geneva, Switzerland.

[2] W. Spirkl and J. Muschaweck, 1992, General Model for Testing Solar Domestic Hot Water Systems, Solar Energy Materials and Solar Cells, 28 (1992), 93-102

[3] A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward, 2005, SUNDIALS: Suite of Nonlinear and Differential/Algebraic Equation Solvers, ACM Transactions on Mathematical Software, 31(3), pp. 363-396

[4] Rui C. Martins, Vitor V. Lopes, António A. Vicente and José A. Teixeira, 2008, Chapter 1.4: Numerical solutions - finite element and finite volume methods in Optimization in Food Engineering, Ferruh Erdogdu (eds) pp 49-108.

[5] R. Siegel and J. R. Howell, 1972, Thermal Radiation Heat Transfer, McGraw-Hill, New York.

[6] John P. Boyd, 2000, Chebyshev and Fourier Spectral Methods, 2nd Ed., Dover Pubs., Inc.

[7] M. Abramowitz and I. A. Stegun, 1972, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover.

[8] E. Keogh, S. Chu, D. Hart and M. Pazzani, 2001, An online algorithm for segmenting time series, ICDM 2001, Proc. IEEE Int. Conf. on Data Mining. 29 Nov 2001 - 02 Dec 2001, San Jose, CA , USA