

# The Role of Learning Methods in the Dynamic Assessment of Power Components Loading Capability

Domenico Villacci, *Member, IEEE*, Gianluca Bontempi, Alfredo Vaccaro, *Member, IEEE*, and Mauro Birattari

**Abstract**—The need for dynamic loading of power components in the deregulated electricity market demands reliable assessment models that should be able to predict the thermal behavior when the load exceeds the nameplate value.

When assessing network load capability, the hot-spot temperature of the components is known to be the most critical factor. The knowledge of the evolution of the hot-spot temperature during overload conditions is essential to evaluate the loss of insulation life and to evaluate the consequent risks of both technical and economical nature. This paper discusses an innovative *grey-box* architecture for integrating physical knowledge modeling (a.k.a. white-box) with machine learning techniques (a.k.a. black-box). In particular, we focus on the problem of forecasting the hot-spot temperature of a mineral-oil-immersed transformer. We perform a set of experiments and we compare the predictions obtained by the grey-, white-, and black-box approaches.

**Index Terms**—Intelligent systems, learning systems, power system monitoring, power transformers protection.

## LIST OF SYMBOLS

$\Theta_A$	Ambient temperature ( $^{\circ}\text{C}$ ).
$\Theta_{TO}$	Top oil temperature ( $^{\circ}\text{C}$ ).
$\Theta_H$	Winding hot-spot temperature ( $^{\circ}\text{C}$ ).
$\hat{\Theta}_H$	Estimated winding hot-spot temperature ( $^{\circ}\text{C}$ ).
$\Theta_{H,R}$	Rated winding hot-spot temperature ( $^{\circ}\text{C}$ ).
$\Theta_{LV}$	Low-voltage winding hot-spot temperature ( $^{\circ}\text{C}$ ).
$\Theta_{MV}$	Medium-voltage winding hot-spot temperature ( $^{\circ}\text{C}$ ).
$\Delta\Theta_H$	Hot-spot temperature rise above top oil ( $^{\circ}\text{C}$ ).
$\Delta\Theta_{TO,U}$	Ultimate top oil temperature rise ( $^{\circ}\text{C}$ ).
$\Delta\Theta_{TO,R}$	Rated top oil temperature rise over ambient ( $^{\circ}\text{C}$ ).
$\Delta\Theta_{H,U}$	Ultimate hot-spot temperature rise over top oil (for a given load current) ( $^{\circ}\text{C}$ ).
$\Delta\Theta_{H,R}$	Rated hot-spot temperature rise over top oil (for rated load current) ( $^{\circ}\text{C}$ ).
$\tau_{TO}$	Top oil rise time constant (h).
$\tau_H$	Hot-spot rise time constant (h).
$I_L$	Load current normalized to rated current (p.u.).

Manuscript received November 8, 2002; revised March 26, 2004. Abstract published on the Internet November 10, 2004.

D. Villacci and A. Vaccaro are with the Power System Research Group, Department of Engineering, University of Sannio, I82100 Benevento, Italy (e-mail: villacci@unisannio.it).

G. Bontempi is with the Departement d'Informatique, Universite Libre de Bruxelles, 1050 Brussels, Belgium (e-mail: gbonte@ulb.ac.be).

M. Birattari is with IRIDIA, Universite Libre de Bruxelles, 1050 Brussels, Belgium (e-mail: mbiro@ulb.ac.be).

Digital Object Identifier 10.1109/TIE.2004.841072

$R$	Ratio of rated-load loss to no-load loss at applicable tap position.
$e_1$	Empirically derived exponent, dependent on the cooling method.
$e_2$	Empirically derived exponent to approximately account for effects of change resistance with change in load.
$t$	Observed time (h).
$L$	Power transformer loss of life (h).
$\Delta t_i$	$i$ th time interval (sample time) (h).
$F_{EQA}$	Aging acceleration factor relative to the $i$ th time interval.
$FAA_i$	Equivalent aging factor for the observed period.

## I. INTRODUCTION

THE safeguard of power components is assuming a major role in the deregulated market of electricity, where a malfunctioning power system could be responsible for serious damage to a large number of *system operators* having access to the shared energy resource. The need for providing a reliable and safe service has historically induced the *asset owner* to adopt a conservative strategy in loading power components. This worst case approach decreases the risk of malfunctioning at the cost of a reduced power transfer capability. As a consequence, the conservative approach appears to be inadequate in the new competitive scenario where the strive for larger profits asks for pushing to the maximum the exploitation of plants. In this scenario, a reliable assessment of load capabilities and an effective management of the associated risks appear to be crucial.

The problem of supplying energy is time varying in nature and demands, therefore, a *dynamic* solution in order to manage risks related to load levels exceeding component's nameplate values, especially in presence of contingencies. In power components an accurate prediction of the evolution of the hot-spot temperature is an essential information to evaluate the risk associated with a given load management policy. This demands the design of a model able to predict the evolution of the hot-spot temperature and the related maximum duration, on the basis of the thermal state, the expected load level, and the forecasted environmental conditions. This model should also exhibit adaptive features, to deal with the intrinsic time-varying phenomena affecting the thermal exchange characteristic of the component (e.g., aging, soil properties, etc.), and low computational times, to comply with the requirements of Energy Management Systems/Distribution Management Systems.

Once defined, the role and the relevant inputs of the thermal model, the power engineer has still to address an important issue: the choice of the most appropriate model structure. As proposed by [1], model structures can be regrouped into the following three main classes:

- **white-box models:** first-principle models that condense the physical insight of the expert; they might contain both physical constants and unknown parameters;
- **black-box models:** a family of linear or nonlinear models whose parameters do not have any physical significance; the goal of black-box modeling is to fit the data rather than gaining insight into the phenomena at hand, and thanks to their data-driven nature these models can be used in an adaptive way to cope with time-varying problems.
- **grey-box models:** an intermediate approach that aims to preserve the best from the previous approaches by integrating knowledge coming from the expert with empirical evidence provided by observations; in particular, the adaptive feature of black-box models is preserved.

White-box models are the classical approach adopted for describing the thermal behavior of power components. They can be divided into two main subclasses: 1) *analytical* and 2) *numerical*. *Analytical thermal models* [2]–[4] are inadequate in presence of changeable climatic conditions and of complex nonlinear phenomena affecting the component thermal dynamic [5]. *Numerical thermal models* [6]–[9] are more detailed and accurate but require complex and time consuming numerical solvers. It follows that they are not exploitable in a dynamic assessment procedure.

Recently, research efforts have been oriented to develop thermal models which should provide high accuracy at low computational costs. In particular, the power systems community started adopting black-box models [10]–[12] thanks to the availability of empirical observations made possible by the advent of hot-spot temperature acquisition modules, based on fiber optical temperature sensing technology [7], [13], [14] or advanced noninvasive identification techniques [15], [16].

To the best of our knowledge, the application of grey-box models has not been so far explored in the literature of thermal modeling of power components. With this paper, we precisely intend to fill this gap. The paper explores dynamic strategies to assess the loading capability of power components and presents experimental results in the case of a distribution transformer. The monitoring of power transformers is an important research subject for several reasons: 1) power transformers are currently subjected to reduced load levels (especially in transmission networks); 2) the performance of power transformers is rapidly improving due to new materials and innovative construction techniques; 3) power transformers (and power cables) are the bottlenecks in networks load capability; and 4) a better load assessment would have a noticeable effect on transmission and distribution networks, especially during emergency conditions.

The paper brings two main contributions to the existing literature. The first is the introduction of a grey-box architecture integrating machine learning techniques [17] with first-principle models. The second is an experimental comparison of white-, black-, and grey-box techniques on short-term and long-term

forecasting of the hot-spot temperature. The black-box techniques considered in this paper are the conventional recursive least-squares (RLS) and some nonlinear learning methods. In particular, we focus on the Lazy Learning (LL) method [18], [19] that differentiates from other nonlinear approaches (like neural networks) on the fact that no functional model of the data is built: for each given query, LL obtains the prediction by: 1) selecting the most relevant samples from the dataset according to a distance metric, and then 2) interpolating the selected points with a simple local approximator. As discussed in Section II-B.1, this procedure can be easily made adaptive to cope with time-varying settings.

The experimental setting proposed in the paper compares white-, black-, and grey-box approaches on the prediction of the winding hot-spot temperature of a 25-kVA mineral-oil-immersed power transformer. The results show the effective role played by adaptive learning methods in addressing the problem of dynamic loading assessment.

The outline of the paper is as follows. A definition of the problem of dynamic loading assessment is given in Section II. In Section III the different modeling approaches to the thermal modeling of a power transformer are introduced. Section IV contains the description of the experiments and a discussion of the results. Conclusions and future work are summarized in Section V.

## II. PROBLEM OF DYNAMIC LOADING CAPABILITY ASSESSMENT

The assessment of the dynamic loading capability of a power component is carried out through two main stages: 1) the identification of the magnitude and the time duration of the electrical load that a power component can support in excess of the design rating and 2) the management of the consequent technical and economical risks. Technical risks are due to mechanical or electrical phenomena (e.g., the aging of the insulation material) or by second order effects specific to a component (e.g., the maximum allowable sag of overhead lines [4] or the generation of free gas bubbles in power transformers [3]). Economic risks are consequent to the damages caused by malfunctioning of the power network (e.g., outage).

An important support to the identification of the risks in loading assessment can be provided by an accurate prediction of the evolution of the hot-spot temperature. This prediction requires the development of a suitable thermal model of the power component in order to predict the evolution of the hot-spot temperature and the associated maximum allowed duration, on the basis of the thermal state, the forecasted environmental conditions and the load level. The prediction can refer to different time horizons ranging from a few minutes (short-term) to several hours (medium-term) or days (long-term load capability estimation). In particular, short-term load capability estimation is essential during system security studies in which the operators evaluate the effect of equipment loadings according to simple ( $n - 1$ ) or more onerous criteria (e.g.,  $n - 2$ ) [20]. As for the medium and long-term load capability estimation, they are oriented to establish an acceptable level of power transfer for a defined time period. This information allows to address effectively several critical issues like the

mitigation of the congestion effects or the exploitation of high price differentials in interconnector circuits trading [20].

In summary, three are the features demanded to a methodology to assess short, medium or long dynamic load capability: 1) high accuracy; 2) adaptive features, in order to manage the intrinsic time-varying phenomena affecting the thermal exchange characteristic of the component (e.g., aging, soil proprieties, etc.); and 3) low computational requirements in order to make hardware implementations feasible. The following section discusses how these characteristics can be addressed by different types of thermal models in the specific case of a power transformer.

### III. DYNAMIC LOADING CAPABILITY ASSESSMENT OF POWER TRANSFORMERS

The variable of interest in the case of a power transformer is the temperature of the hottest part of the windings. An accurate prediction of this quantity allows an accurate estimation of the insulation state and the components' residual thermal life. Unfortunately, accurate hot-spot temperature evaluation is a difficult task as stressed by the numerous studies carried out on the subject [5], [10], [12]. Difficulties arise due to the heat transfer process which is distributed over many complex surfaces made up of different materials. This makes the mathematical description of the temperature distribution in the transformer very complex. This section reviews into detail three main approaches (white, black, and grey) that can be pursued in order to address the problem of modeling the thermal dynamic of a power transformer.

#### A. White-Box Modeling Approach

The instantaneous evolution of the winding hot-spot temperature at the top or in the center of the high or low voltage winding of a power transformer can be estimated by solving the analytical model described in [3], and updated in [2]. The simplifying assumptions adopted in the formulation of such a model are: the oil temperature profile inside the winding increases linearly from bottom to top; the difference between the winding temperature and the oil temperature is constant along the winding; the ambient temperature drives the oil temperature up and down with the same time constant as the winding temperature does; the solar flux incidence is neglected. Such assumptions led to the adoption of the temperature profile inside the transformer depicted in Fig. 1.

Once these assumptions are made, the hot-spot temperature  $\Theta_H$  can be calculated as the sum of two components, the top oil temperature  $\Theta_{TO}$  and the hot-spot rise above top oil temperature  $\Delta\Theta_H$  as expressed in the following equation:

$$\Theta_H = \Theta_{TO} + \Delta\Theta_H. \quad (1)$$

As reported by [2] the evolution of such variables can be estimated by the following physical model:

$$\begin{cases} \tau_{TO} \frac{d\Theta_{TO}}{dt} = [\Delta\Theta_{TO,U} + \Theta_A] - \Theta_{TO} \\ \tau_H \frac{d\Delta\Theta_H}{dt} = \Delta\Theta_{H,U} - \Delta\Theta_H \\ \Delta\Theta_{TO,U} = \Delta\Theta_{TO,R} \left[ \frac{I_L^2 R + 1}{R + 1} \right]^{e_2} \\ \Delta\Theta_{H,U} = \Delta\Theta_{H,R} I_L^{2e_1} \end{cases} \quad (2)$$

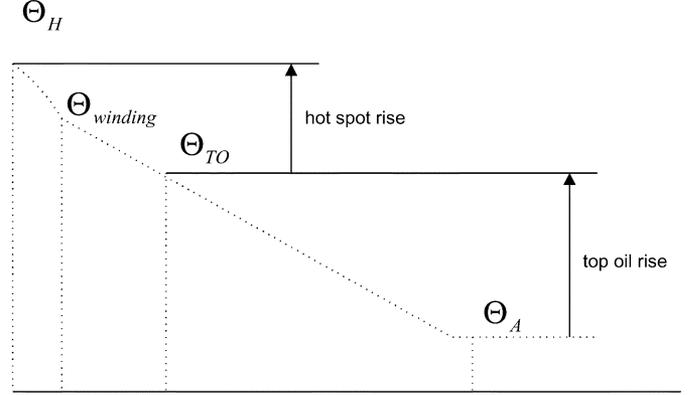


Fig. 1. Temperature profile assumed in the IEEE loading guide.

This simplified equivalent model requires the setting of some specific transformer parameters, like  $\tau_{TO}$  which represents the top oil temperature time constant. These parameters can vary considerably from one transformer to another, making then the accuracy of model (2) extremely sensitive to parameter variations and overload conditions [5], [10]. Adopting inaccurate thermal models can induce a substantial error in determining the real-time transformer load capability ratings: The loss of life is exponential in the hot-spot temperature and inaccurate predictions of the latter have dramatic effects on the accuracy of the former. Additional warning should come from the fact that for large transformers the computation of the load capability becomes more complex and inaccurate due to the leakage flux that causes heating of structural parts with a consequent intensification of gas-bubble formation [3]. This inaccuracy holds in spite of the corrective terms that could be included in the model (2) in order to take into account the cooling effects of forced oil circulation [5].

#### B. Black-Box Modeling Approach

The general black-box approach to model an input/output phenomenon (with the scalar  $y$  as output, and the vector  $\varphi$  as input) relies on the availability of a collection of observed pairs  $\{[y(t), \varphi(t)]; t = 1, \dots, N\}$  typically called the *training set*. The prediction problem consists in predicting the value of the output  $y(N + 1)$  when the value of the input is  $\varphi(N + 1)$ . Learning methods [17] aim to find a suitable function  $f(\cdot)$  such that the output variable can be accurately represented by a model in the form

$$y(t) = f(\varphi(t)) + v(t) \quad (3)$$

where  $v(t)$  is usually thought of as the term including modeling error, disturbances and noise. In the machine learning community, the problem (3) is often referred to as a *supervised learning* problem. If conventional assumptions of normality and whiteness are made on  $v(t)$ , it follows that a reliable prediction of the output, given the input, is returned by

$$\hat{y}(t) = f(\varphi(t)) \quad (4)$$

Traditional approaches address linear formulation of the model  $f(\cdot)$

$$\hat{y}(t) = \theta^T \varphi(t) + v(t) \quad (5)$$

where the vector of parameters  $\theta$  is estimated on the basis of the observed dataset by using conventional least-squares techniques. Recent advances in machine learning and data mining have proposed a set of powerful approximators of nonlinear relationships  $f(\cdot)$  [17], as discussed in the following section.

The formulation in (3) covers a number of prediction problems but especially applies to the generic discrete-time nonlinear autoregressive exogenous (NARX) system identification setting [17] where the input vector takes the form

$$\varphi(t) = [y(t-k), \dots, y(t-k-n+1), u(t-1), \dots, u(t-m)]^T \quad (6)$$

the constant  $k \geq 1$  defines the time horizon of the prediction problem, and  $u(t)$  denotes the exogenous input signal of the dynamical system. Note that the number  $n$  of lagged outputs and the number  $m$  of lagged control actions are properly set in order to describe accurately the dynamic of the system.

In our transformer problem, the output variable  $y$  is the hot-spot temperature  $\Theta_H$ , while the exogenous inputs are the current load  $I_L$  and the environmental conditions  $A$ . In fact, the practical experiments neglect the environmental conditions  $A$ , as they are supposed to remain practically constant all along the experiences. Their effect is incorporated in the noise term  $v$ . This is the resulting NARX form of our black-box predictor

$$\hat{\Theta}_H(t) = f(\Theta_H(t-k), \Theta_H(t-k+1), \dots, \Theta_H(t-2k+1), I_L(t-1), I_L(t-2), \dots, I_L(t-k)) \quad (7)$$

where a single step corresponds to 5 min and the horizon  $k \geq 1$  is expected to range over the interval [1, 12] (i.e., from 5 min to 1 h). This model aims at predicting the hot-spot temperature at time  $t$ , when the last available measure has been collected at time  $t-k$ . In order to perform the prediction the model requires the window of latest  $k$  values of  $\Theta_H$  (i.e.,  $n = k$ ) and the load profile of  $I_L$  up to time  $t-1$  (i.e.,  $m = k$ ). The choice of the orders  $m$  and  $n$  has been suggested by a procedure of cross-validated feature selection [17] performed on the available experimental data.

In the following, we refer to model (7) as the *short-term black-box predictor* of the hot-spot temperature. Two are the remaining issues still to be addressed in the definition of the short-term predictor. The first is to define the structure of the unknown function  $f(\cdot)$ , the second is to define an adaptive procedure able to update sequentially the estimate of  $f(\cdot)$  once new observations are available.

As far as linear relationships  $f(\cdot)$  are taken into account, the identification procedure is well established (least-squares algorithm) and the recursive procedure is well known (RLS: recursive least squares) [21]. The problem becomes harder in terms of prediction accuracy and computational requirements when we move to the nonlinear case.

1) *LL for Black-Box Modeling*: In the Machine Learning literature [22], different criteria have been proposed for classifying

learning methods. Those of interest in this paper are based on the dichotomies *global/local*, *lazy/eager*, and *linear/nonlinear*. Few examples will be sufficient for clarifying these concepts. Neural networks (NN) are classical instances of the global, eager, and nonlinear approach: NN are global in the sense that a single representation covers the whole input space. They are eager in the sense that the examples are used for tuning the network and then they are discarded without waiting for any query. Finally, NN are nonlinear in the sense that the relation between the weights and the output is nonlinear. The classical linear regression used by statisticians is an example of global, eager, and linear approach. The LL method that we discuss in this paper is a lazy and local approach: a dataset of examples is stored, and any processing is deferred until an explicit request for a prediction is received. When this happens, the dataset is searched for examples falling in a neighborhood of the query point. Such examples are used for identifying a local model that is then evaluated in the query point to return a prediction. The local model is then discarded and the procedure is repeated from scratch for subsequent queries.

The major appeal of LL is precisely its divide-and-conquer nature: LL reduces a complex and nonlinear modeling problem into a sequence of easily manageable local linear problems, one for each query. This allows to exploit, on a local basis, the whole range of linear identification and validation techniques which are fast, reliable, and come with a wealth of theoretical analyzes, justifications, and guarantees.

The LL procedure essentially consists of the following steps [23], [24].

- Associate at each sample  $(\varphi_i, y_i)$  a weight factor  $w_i$  in function of the distance  $d(\varphi_i, \varphi_q)$  and generate the matrices  $Z = W\Phi$  and  $v = Wy$  where  $W$  is a diagonal matrix having diagonal elements  $W_{ii} = w_i$ .
- Solve, using a number  $\hat{k}$  of neighbors of  $\varphi_q$ , a linear locally weighted regression problem

$$(Z^T Z)\beta = Z^T v \quad (8)$$

where  $\beta$  is the vector descriptive of the first order polynomial used as local approximator.

- Calculate the prediction by evaluating the model (8) in the query point  $\varphi_q$  as

$$\hat{y}_q = \varphi_q^T (Z^T Z)^{-1} Z^T v. \quad (9)$$

This algorithm demands the choice of the number  $\hat{k}$  of neighbors of  $\varphi_q$  to be used in the local regression fit and the tuning of the local model parameters. The black- and the grey-box approaches proposed in this paper are based on a version of LL that, on a query-by-query basis, tunes the number  $\hat{k}$  of neighbors [23], [24]. This is done by solving in a recursive fashion the (8). The idea consists in starting from a minimum number of neighbors and recursively adding neighbors until the predicted performance of the corresponding local approximation decays significantly or a maximum number of examples is reached [23]. More details on the LL technique and its applications can

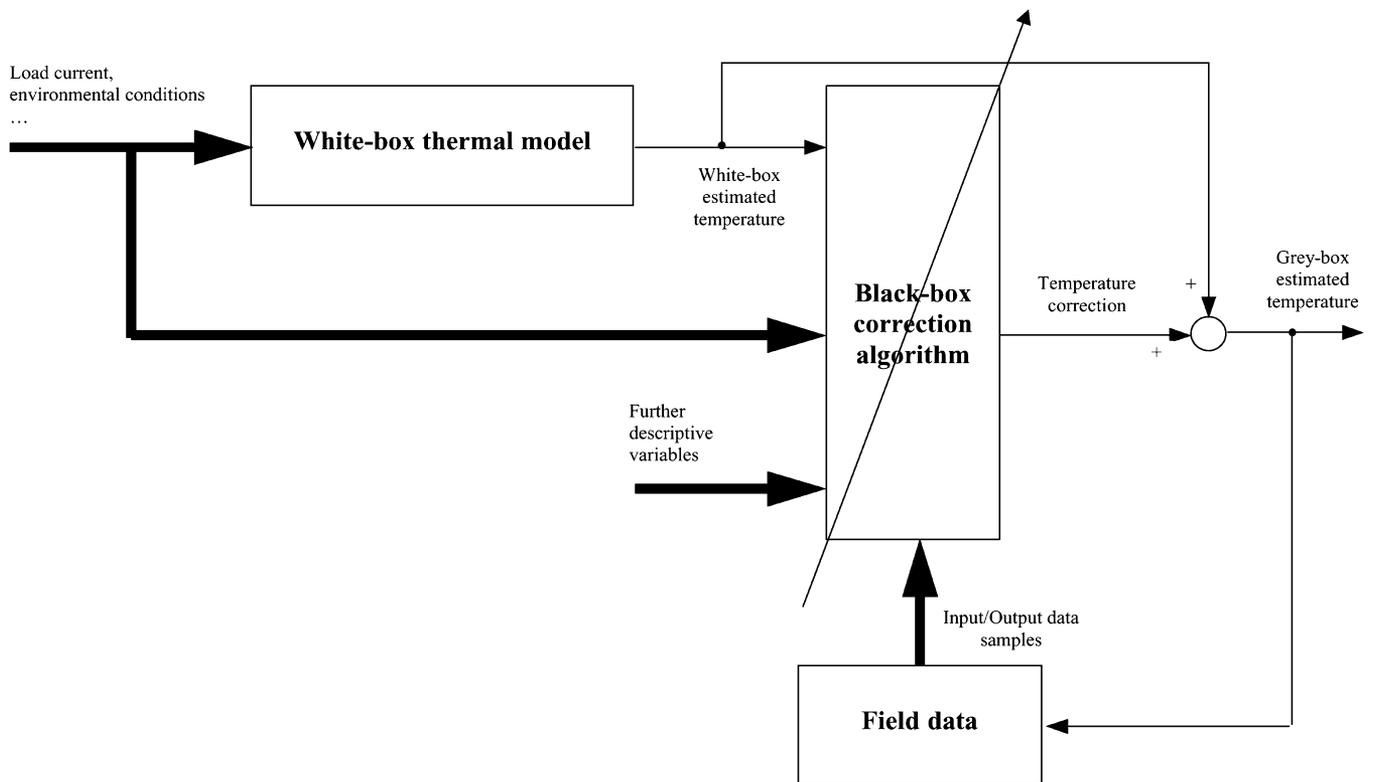


Fig. 2. Grey-box modeling architecture.

be found in [19], [24], [26], and [27]. Also, a comparative analysis of LL and neuro-fuzzy inference systems for modeling is presented in [6] and [25].

This paper illustrates and validates the use of LL in the specific problem of power transformer thermal modeling. The LL algorithm exhibits a number of features that match the characteristics of the hot-spot temperature prediction problem.

- **The reduced number of assumptions:** LL assumes no a priori knowledge on the process underlying the data. For example, it makes no assumption on the existence of a global function describing the data and no assumptions on the properties of the noise. The only available information is represented by a finite set of input/output observations. This feature is particularly relevant in real datasets where problems of missing features, nonstationarity and measurement errors make appealing a data-driven and assumption-free approach.
- **Online learning capability:** LL can easily deal with online learning tasks where the number of training samples increases with time. In this case, LL simply adds new points to the dataset: unlike neural networks, LL does not need time-consuming re-training when new data become available.
- **Modeling nonstationarity:** LL can deal with time-varying configurations where the stochastic process underlying the data is nonstationary. In this case, it is sufficient to interpret the notion of neighborhood not in a spatial way but both in a spatial and temporal sense. For each query point, the neighbors are no more the samples that have similar inputs but the ones that both have similar inputs and have

been collected recently in time. Therefore, the time variable becomes a further precious feature to consider for accurate prediction.

These considerations motivate the adoption of the LL algorithm as an adaptive nonlinear black-box estimator and as an adaptive correction estimator in the grey-box architecture (Section III-C). Its accuracy is assessed in Section IV.

### C. Grey-Box Modeling Approach

The grey-box modeling approach pursues the effective integration of the physical knowledge available about the system with the additional information retrievable from experimental measurements. This paper proposes a grey-box architecture to improve the accuracy of white-box models in assessing the loading capability of the transformer. The basic idea is to combine the prediction returned by the physical transformer thermal model with adaptive black-box correction algorithms. The proposed grey-box architecture is sketched in Fig. 2.

We expect this architecture to be reliable both in terms of robustness and speed since it combines a fast built-in physical model with an adaptive dynamic corrector. The corrector model is sequentially updated, in order to adapt the whole architecture to “new” operating conditions. A way to detect a new condition is checking regularly the prediction accuracy on fresh measurements: a prediction error over a fixed threshold could be used to trigger the dataset update. It is worth noting also that this structure, by avoiding recursive feedbacks, should in principle also prevent stability problems and the consequent explosion of the error (e.g., like in recurrent neural architectures).

The design of the proposed grey-box scheme relies on two main components: 1) a reliable white-box physical model describing the transformer and 2) an effective adaptive model able to predict the correction on the basis of measurements.

We consider the IEEE model described in (2) and LL for playing the role of the physical model and of the adaptive corrector, respectively. We expect that their integration should ensure an high accuracy in the thermal estimation process providing, at the same time, the capability to adapt to the inherent time-varying phenomena characterizing the thermal dynamic. Another nonnegligible factor on behalf of this combination relates to the advantageous computational requirements in view of a hardware implementation. Low computational cost is guaranteed by the fact that: 1) the white-box is a simplified thermal model and 2) the LL requires neither a heavy structural identification phase nor a long updating process.

1) *The Identification of the Corrective Module:* The correction module is the black-box component of the grey-box architecture. This module takes as input the information coming from the physical model and the measured information coming from the observations. The expected output is a correction of the prediction returned by the white-box model.

Let  $\Theta_H(t)$  be the real hot-spot temperature at time  $t$  and  $\Theta_m(t)$  the prediction of  $\Theta_H(t)$  returned by the physical model. We denote by  $E_H(t) = \Theta_H(t) - \Theta_m(t)$  the white-box prediction error. Once  $\Theta_H(t)$  and  $\Theta_m(t)$  are available, the quantity  $E_H(t)$  can be easily measured and stored in the observation database. It follows that a black-box predictor can be used to predict the value of  $E_H(t)$ , given a set of inputs. We propose the following black-box structure to address the problem of short-term forecasting:

$$\hat{E}_H(t) = f(\Theta_H(t-k), \Theta_m(t), I_L(t-1)) \quad (10)$$

where the horizon ranges over the interval [1, 12] (or equivalently in the range 5 min, 1 h). Note that the accessibility of the physical prediction  $\Theta_m(t)$  allows the use of a reduced set of input features with respect to the black-box formulation in (7). While the black-box approach in (7) requires a window of  $k$  past values, the grey-box approach can find equivalent information in the physical quantity  $\Theta_m(t)$  returned by the white-box predictor. This shows clearly the benefit of the semi-physical grey-box approach with respect to a pure black-box approach.

Experimental results show that the short-term correction model returns an accurate prediction only for reduced values of  $k$ . Increasing the gap between the last measure of  $\Theta_H(t)$  and the prediction has as inevitable consequence the explosion of the prediction error. This phenomenon rises the need for addressing long-term assessment with a different correcting model structure. Hence, we propose the following long-term formulation:

$$\hat{E}_H(t) = f(\Theta_m(t), I_L(t-1)). \quad (11)$$

Experiments show that this corrective structure allows a more robust behavior than the correction (10) on middle (more than 1 h), long (one day), and very long horizons (two days). Since this model structure is equivalent to the structure in (10) apart from the removal of the measure of  $\Theta_H(t)$ , some considerations

have to be made. We maintain that the better performance of (11) with respect to (10) is due to a reduced complexity of the corrector which makes it less prone to errors on long horizons. At the same time, what seems a loss of information with respect to (10) (i.e., the absence of  $\Theta_H(t)$ ) is only apparent since the variable  $\Theta_H(t)$  re-enters the model as initial condition of the computation of  $\Theta_m(t)$ .

#### IV. EXPERIMENTAL SETTING

*The measurement station (Fig. 3):* It is formed of a set of fiber-optical-based sensors that measures: 1) the hot-spot temperature of the medium-voltage and low-voltage windings and 2) the top oil temperature [10], [15], [16]. The transformer main characteristics are reflected in Table I. In order to measure the load current a Hall-effect current transducer is used. The ambient temperature is monitored through a digital thermometer located far enough away from the power transformer so that the impact of heat dissipated from the transformer on the ambient temperature can be neglected. All sensors are interfaced with a data acquisition unit, which is used also for controlling the variac tap-changer. A data logging system records the temperature registered by each sensor at 5-min intervals. The measurement session consisted in imposing several realistic loading current profiles and in acquiring the load current, the transformer top oil temperature, the weather conditions, and the corresponding winding hot-spot temperature.

*Data collection:* The data were recorded during 24 h by loading the transformer according to the profile reported in Fig. 4(a). This profile is expected to be representative of the different operating conditions of the transformer as it contains both situations of normal load and an overload condition of about 2 h (around the seventh hour of functioning). The related hot-spot behavior is reported in Fig. 4(b).

The training set, resulting from the from the measurement procedure, is composed of  $N = 290$  input/output pairs<sup>1</sup>.

*Validation procedure:* The prediction experimental session adopts a *training-and-test* procedure [22] in order to have an accurate assessment of the accuracy of the different modeling approaches. As measures of accuracy we use the root mean square error (RMSE) and the peak error (PE)  $PE = \hat{\Theta}_H(t) - \Theta_H(t)|_{\max(\Theta_H(t))}$  at the winding hot-spot temperature. The PE quantity represents an essential figure of merit in the evaluation of the prediction accuracy since it describes the suitability of the prediction algorithm to identify the hot-spot evolution during overloads. Overloads are indeed very critical situations where an extremely high accuracy is required.

The accuracy of the modeling approaches is tested over two different test sets, featuring two different load current profiles: the first without overload [Fig. 5(a)] and the second with a 4-h overload condition of 1.3 p.u. [Fig. 5(b)].

<sup>1</sup>This dataset is indeed a subset of the original set, and it has been obtained by a preprocessing phase which removed those examples which are well modeled or redundant according to a given index of performance. The rationale for this subsampling is to store only those sample points that are relevant for the description of the system dynamics, discarding the ones that are redundant or less meaningful. In particular, we adopt a tolerance threshold of 0.5 °C for load current higher than 0.8 p.u. (per unit) and 2 °C otherwise.

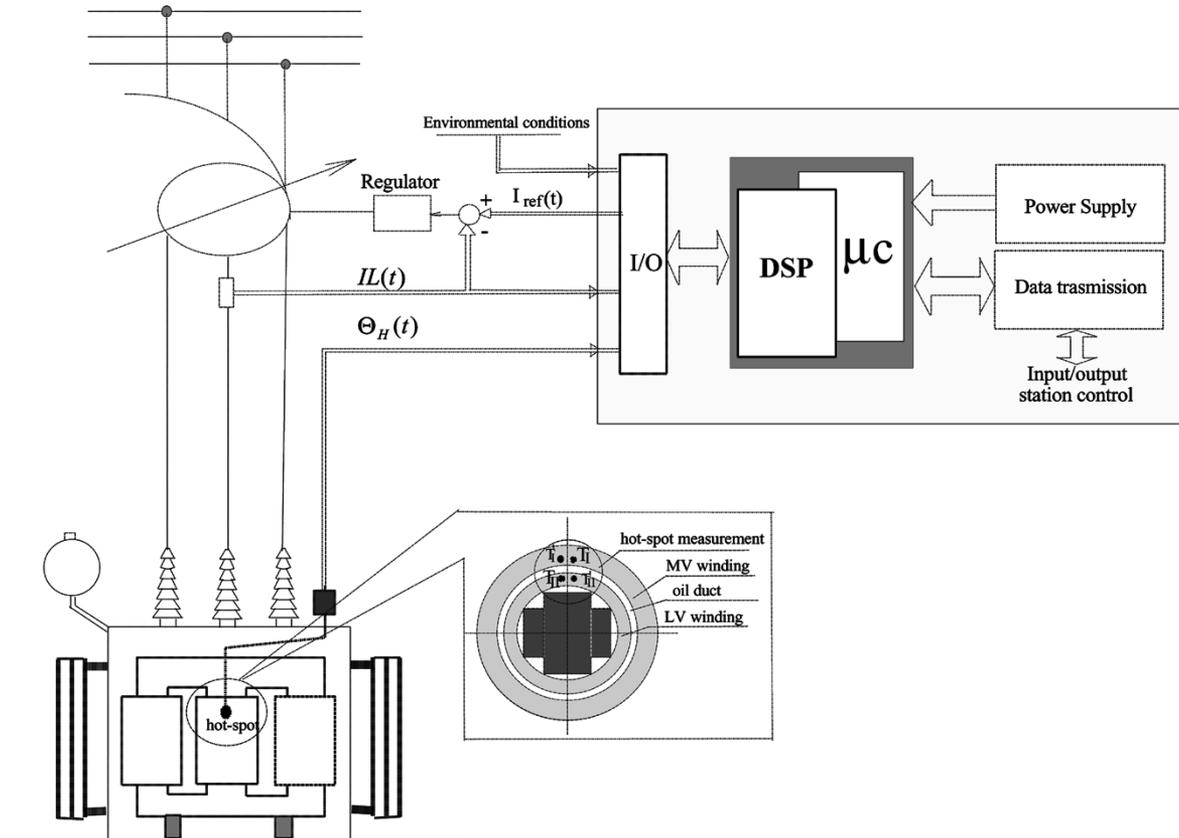


Fig. 3. Power transformer.

TABLE I  
TRANSFORMER CHARACTERISTICS

Nameplate Rating	25 kVA
V1/V2	10kV/380V
Iron losses	195 W
Copper losses (Full load)	776 W
Top Oil Temperature Rise at Full Load	70.1 °C <sup>(*)</sup>
Weight of Core and Coil Assembly	136 kg
Weight of Oil	62 kg
Total Weight	310 kg
Length, Width and Height of Tank	64x16x80cm
Type of Cooling	ONAN
Factory/Year	MACE/87

(\*) values reported to 75°C  
(<sup>o</sup>) with environmental temperature of 21 °C

All computation tasks were performed by using Matlab, version 6<sup>2</sup>. Standard Matlab toolboxes (neural network toolbox) and our own LL code [28] were employed in the experiments.

*Short-term forecasting results:* A linear RLS algorithm and a nonlinear LL<sup>3</sup> algorithm are used to approximate the unknown relationship  $f(\cdot)$ . We explore different prediction horizons ranging between 5 min. and 1 hour. For each time horizon, we report the accuracy of the prediction models assessed over the two validation datasets. An adaptation mechanism is used to

<sup>2</sup><http://www.mathworks.com>

<sup>3</sup>Hereafter, by LL we refer to an LL algorithm which adopts the model combination paradigm ( $b = 10$  in (11)) and where the number of neighbors is allowed to range in the interval [50], [70].

update sequentially the knowledge base with new samples when the prediction error is detected to be worse than the threshold discussed in footnote 1. The results of the black-box model are reported in Fig. 6(a) (RMSE versus horizon). Note that the RMSE error in Fig. 6 is averaged over the two validation sets.

The results of the grey-box model are reported in Fig. 6(b) (RMSE versus horizon).

Looking at the figures we remark that the accuracy deteriorates for larger prediction horizons. This justifies the adoption of specific methods for the long-term forecasting. Also, the analysis of the results reveals that LL and RLS ensure a good level of accuracy both in the black-box and the grey-box architecture. As far as the comparison grey/black is concerned, it appears that in this case the performance of grey-box models does not significantly outperform black-box methodologies. A substantial equivalence in terms of prediction accuracy should then lead us to favor less complex methodologies. On this matter, it appears that the grey-box requires a simpler identification model (i.e., less input features). In terms of robustness this should guarantee a better behavior in front of unexpected configurations.

As far as a comparison LL-RLS is concerned, a considerable difference emerges with reference to the PE evolution (figures not available for lack of space). In this context the experimental results show that LL gives better accuracy than RLS. This is probably due to a specific nonlinear effects appearing in over-load regimes.

*Long-term forecasting:* The application of the short-term models to a long-term forecasting problem does not guarantee an acceptable performance in terms of accuracy (Fig. 7).

TABLE II  
LONG-TERM FORECASTING ACCURACY

	Validation set 1		Validation set 2	
	PE	RMSE	PE	RMSE
White(IEEE)	-3.27	7.65	5.58	6.25
Grey (LL)	0.26	1.58	1.47	1.30
Grey(RLS)	-1.47	3.95	11.52	4.75

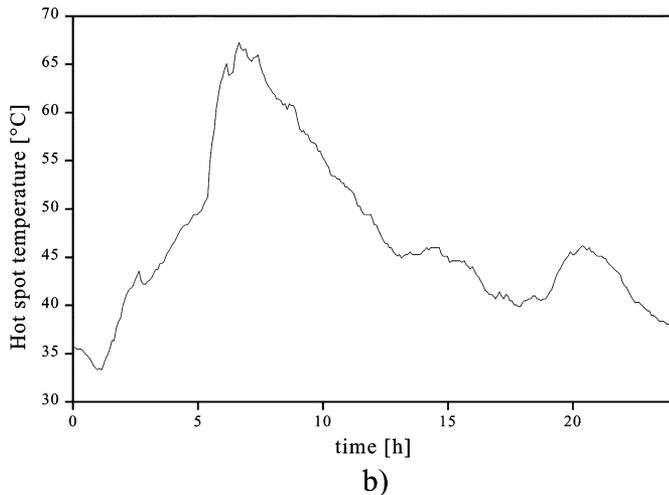
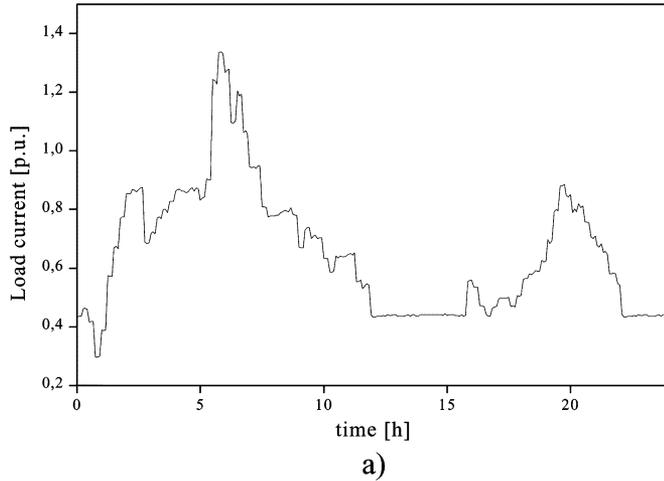


Fig. 4. Training set. (a) Current load profile. (b) Hot-spot behavior.

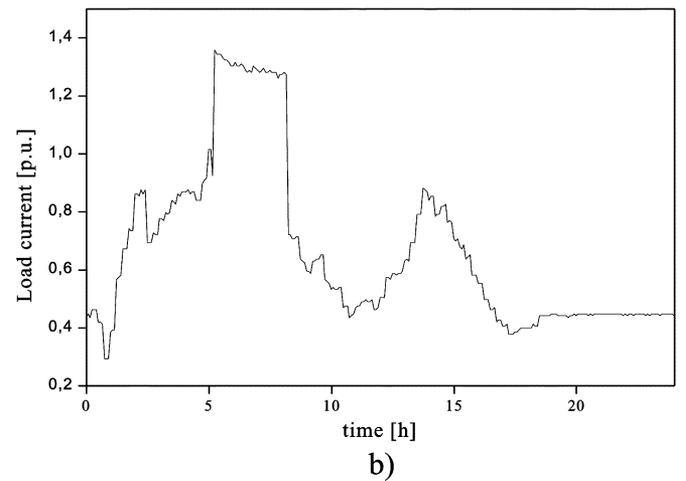
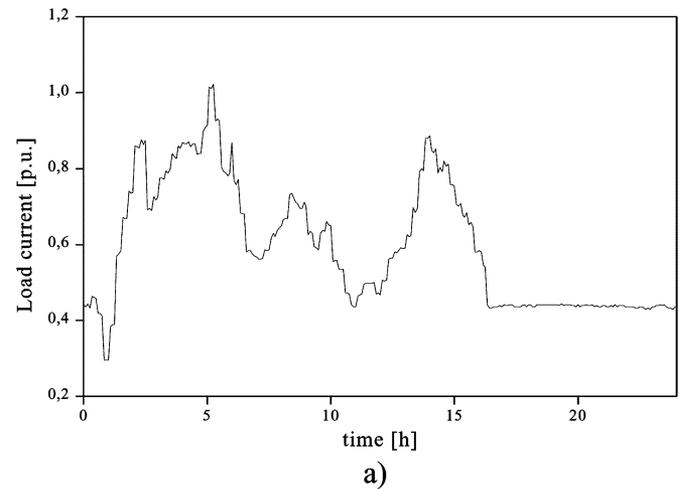


Fig. 5. Load current profiles. (a) First test set. (b) Second test set.

This suggests the adoption of alternative models to address this specific task. We present here the results obtained by using the white-box and the grey-box architectures (Section III-C).

Note that the white-box model is based on the IEEE thermal model where the transformer characteristic parameters are

$$R = 4, \Delta\Theta_{H,R} = 5^\circ\text{C} \quad e_2 = 0.8 \quad \Delta\Theta_{TO,R} = 54^\circ\text{C} \\ e_1 = 0.8 \quad \tau_{TO} = 3\text{h} \quad \Theta_{H,R} = 78^\circ\text{C} \quad \tau_H = 0.1\text{h}$$

Fig. 8(a) reports three curves for the first test set: the measured hot-spot temperature for the first test set, the predictions returned by the white-box (IEEE) model, and the one returned by the grey-box model where the correcting module is implemented by a LL algorithm. Fig. 8(b) reports the same curves for the second test set.

Looking at the PE error (i.e., the error at the maximum temperature) we see that the white-box model underestimates the

temperature, then providing a conservative safety factor to protect the power transformer. A much lower PE is returned by the grey-box models.

Table II summarizes the experimental results obtained so far.

Table II shows that, in terms of RMSE, the grey-box architecture outperforms significantly the white-box approach especially in presence of severe overload. It is worth noting that although the two grey-box models are trained using the same dataset, the LL-based correction algorithm exhibits a better accuracy in the presence of overload conditions.

*Dynamical load capability assessment:* The availability of a predictive model of the hot-spot temperature is a necessary precondition to perform the dynamical load capability assessment procedure. This is carried out, once the real transformer thermal state and the forecasted environmental conditions are given, by solving iteratively the short or long predictive models for a set

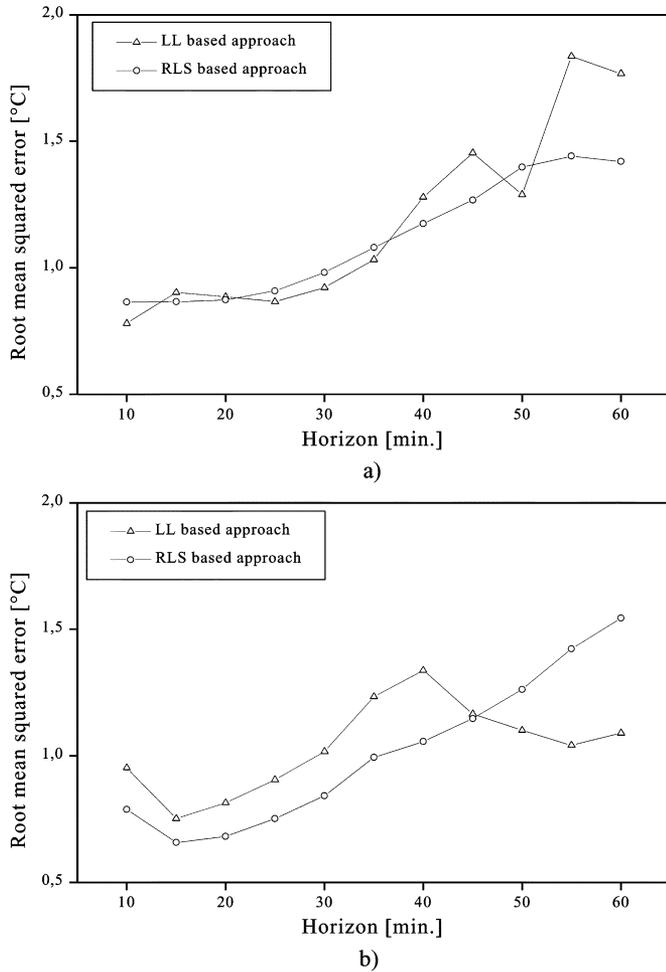


Fig. 6. Short-term forecasting. The figure reports the RMSE versus the time horizon. (a) Black-box linear (RLS) and black-box nonlinear (LL) prediction. (b) Grey-box linear (RLS) and grey-box nonlinear (LL) prediction.

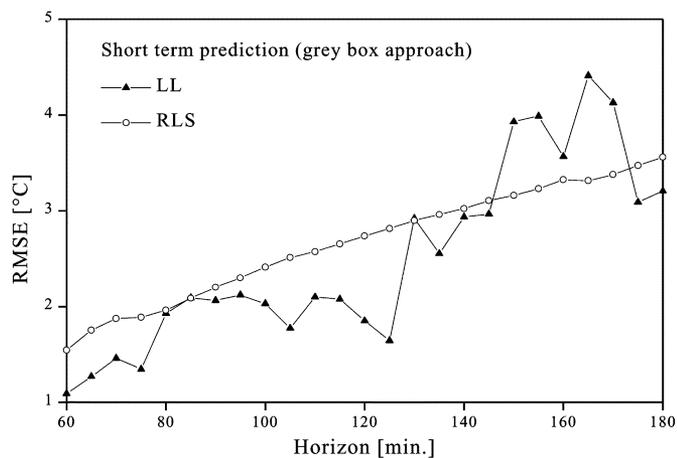


Fig. 7. Long-term forecasting. RMSE accuracy versus horizon of two short-term prediction models.

of loads comprised between 1 and 2 p.u.. The corresponding hot-spot and aging profiles allow to identify the maximum allowable duration for each load. The resulting diagram is called the *load capability curve*. Dynamic load capability curves make possible technical and economical considerations

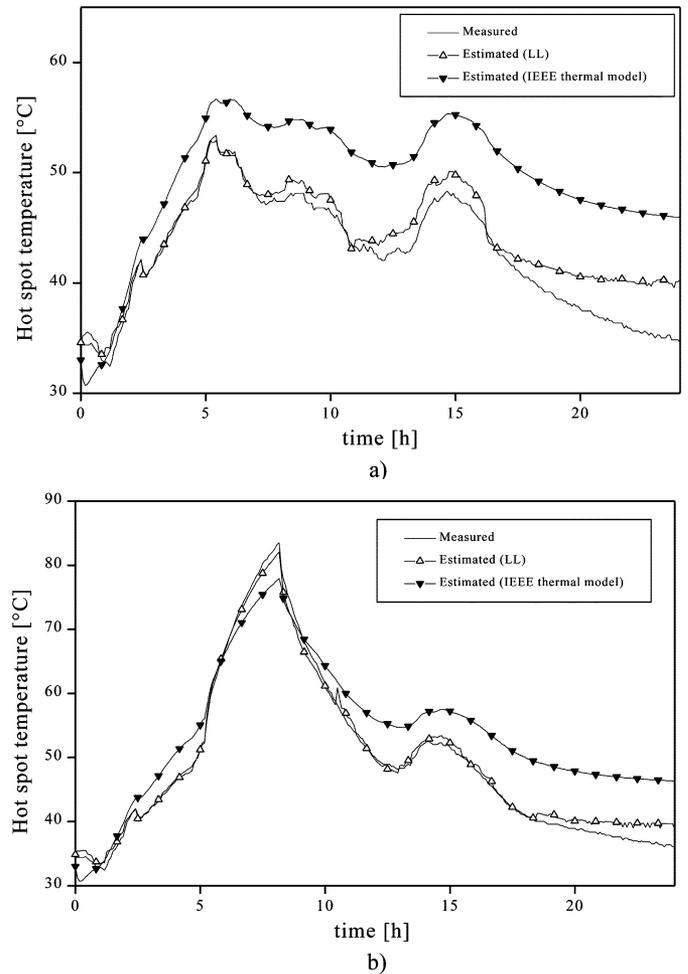


Fig. 8. Long-term forecasting. The figure reports the real hot spot and the predictions returned by the white-box (IEEE) and the grey-box nonlinear (LL) models. (a) First test test. (b) Second test test.

about the load management. For example it allows to fix a set of thresholds for the winding hot-spot temperature, for the insulation loss of life and for the temperature of bushings.

Fig. 9 reports the load capability curves identified starting from two different thermal states of the transformer (cold state and hot state) and adopting as technical constraint a maximum hot-spot threshold of 90 °C. Fig. 9(a) and (b) show the curve obtained by the IEEE model and the LL grey-box model. Again, the gap of the two curves puts into evidence the important side effect of an inaccurate prediction of the hot-spot temperature.

## V. CONCLUSION AND FUTURE DEVELOPMENTS

The changing scenario in the energy market asks the asset owners for new loading strategies in order to attain high profits without losing reliability and security. This work shed light on an innovative approach to loading assessment which aims at integrating consolidated knowledge about the components with innovative techniques of data analysis. Although the experimental study was limited to a power transformer, we expect that the grey-box architecture could exhibit similar promising

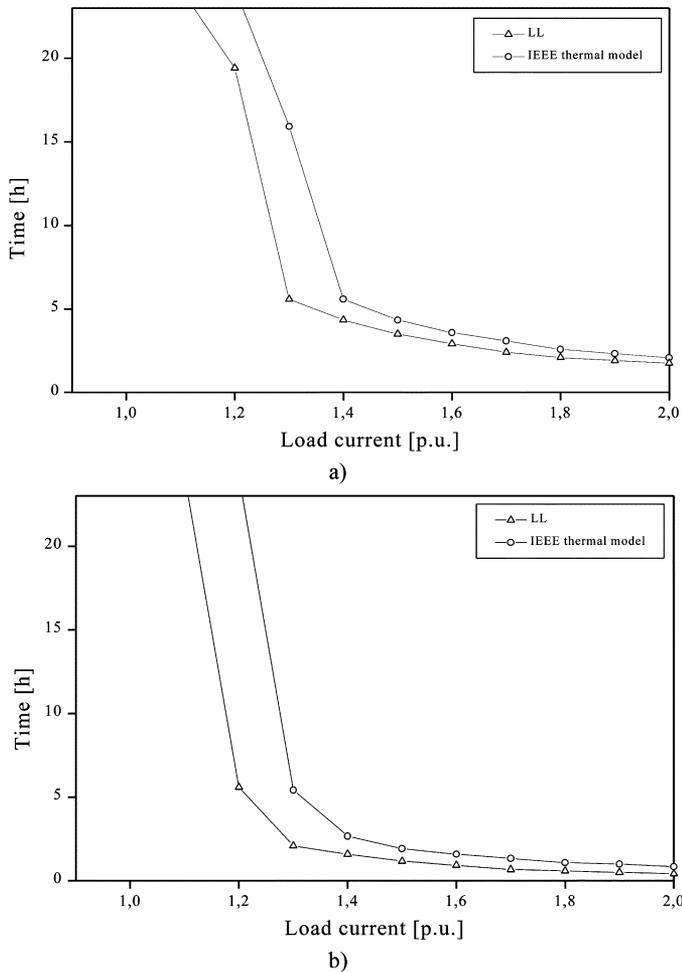


Fig. 9. Long-term forecasting. Load capability curve returned by the white-box and the grey-box (LL) models. (a) Hot-spot temperature = 35° (cold state). (b) Hot-spot temperature = 75° (hot state).

results for other power components, like power electronics devices, power cables, and overhead lines.

Further work will address also an hardware implementation of the modeling architectures on a microcontroller-based unit in order to develop advanced overloading protection units. The availability of sophisticated micro controller, including complete web based functionalities opens the ways to interesting future scenarios like Energy Management Systems and Distribution Management Systems featuring advanced functionalities like remote control, monitoring and assessment (e-assessment). Moreover, a diffused adoption of Web-based micro controller units, equipped with reliable predictive models, could ensure a capillary assessment of the overall network capability raising the operative margins especially during emergency conditions.

## REFERENCES

- [1] P. Lindskog and L. Ljung, "Tools for semiphysical modeling," *Int. J. Adapt. Control Signal Process.*, vol. 9, no. 6, pp. 509–523, Dec. 1995.
- [2] Working Group K3 (Transformer Thermal Overload Protection), "Adaptive transformer thermal overload protection," *IEEE Trans. Power Del.*, vol. 16, no. 4, pp. 516–521, Oct. 2001.
- [3] *Guide for Loading Mineral-Oil-Immersed Transformers*, IEEE Std. C57.91-1995.

- [4] *Calculation of Bare Overhead Conductor Temperatures*, IEEE Std. 783-1993.
- [5] V. Galdi, L. Ippolito, A. Piccolo, and A. Vaccaro, "Parameters identification of power transformers thermal model via genetic algorithms," *Elect. Power Syst. Res.*, vol. 60, no. 2, pp. 107–112, Jan. 2001.
- [6] E. Bertolissi, M. Birattari, G. Bontempi, A. Duchateau, and H. Bersini, "Data-driven techniques for direct adaptive control: The lazy and the fuzzy approach," *Fuzzy Sets Syst.*, vol. 28, no. 1, pp. 3–14, 2001.
- [7] J. S. Lyall, G. Nourbakhsh, and H. C. Zhao, "Underground power cable environment on line monitoring and analysis," in *Proc. 2000 IEEE Power Engineering Soc. Summer Meeting*, Seattle, WA, 2000, pp. 457–462.
- [8] M. A. Hanna, A. Y. Chikhani, and M. M. A. Salama, "Thermal analysis of power systems in a trench in multi-layered soil," *IEEE Trans. Power Del.*, vol. 13, no. 2, pp. 304–309, Apr. 1998.
- [9] G. Buonanno, A. Carotenuto, M. Dell'Isola, and D. Villacci, "The effect of radiative and convective heat transfer on thermal transient in power cables," in *Proc. IEE—Generation, Transmission, Distrib.*, vol. 142, Jul. 1995, pp. 436–444.
- [10] P. Daponte, D. Grimaldi, A. Piccolo, and D. Villacci, "A neural diagnostic system for the monitoring of transformer heating," *Measurement*, vol. 18, no. 1, pp. 35–46, 1996.
- [11] A. Losi, F. Rossi, and D. Villacci, "A new method for the full exploitation of electrical lines," in *Proc. Power System Conf.*, Tehran, Iran, 1993, pp. 416–426.
- [12] W. Van der Veken, J. Declercq, M. Baelmans, and S. Van Mileghem, "New perspectives to overloading with accurate modeling of thermal transients in oil-immersed power transformers," in *Proc. IEEE-PES Transmission and Distribution Conf. and Exhibition*, Atlanta, GA, 2001, pp. 147–152.
- [13] G. Nokes, "Optimising power transmission and distribution networks using optical fiber distributed temperature sensing systems," *Power Eng. J.*, vol. 13, no. 6, pp. 291–296, Dec. 1999.
- [14] J. A. Williams, J. H. Cooper, T. J. Rodenbaugh, G. L. Smith, and F. Rorabaugh, "Increasing cable rating by distributed fiber optic temperature monitoring and ampacity analysis," in *Proc. IEEE-PES Transmission and Distribution Conf.*, New Orleans, LA, 1999, pp. 128–134.
- [15] V. Galdi, L. Ippolito, A. Piccolo, and A. Vaccaro, "Neural diagnostic system for transformer thermal overload protection," in *Proc. IEE—Elect. Power Appl.*, vol. 147, Sep. 2000, pp. 415–421.
- [16] —, "Application of local learning techniques for power transformer thermal overload protection," in *Proc. IEE—Elect. Power Appl.*, vol. 148, Mar. 2001, pp. 163–170.
- [17] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning. Data Mining, Inference and Prediction*. Berlin, Germany: Springer-Verlag, 2001.
- [18] C. G. Atkeson, A. W. Moore, and S. Schaal, "Locally weighted learning," *Artif. Intell. Rev.*, vol. 11, no. 1–5, pp. 11–73, 1997.
- [19] G. Bontempi, M. Birattari, and H. Bersini, "A model selection approach for local learning," *Artif. Intell. Commun.*, vol. 13, no. 1, pp. 41–48, 2000.
- [20] Study Committee 23-CIGRE, "Dynamic loading of transmission equipment," *Electra*, no. 202, pp. 63–73, Jun. 2002.
- [21] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [22] T. M. Mitchell, *Machine Learning*. New York: McGraw-Hill, 1997.
- [23] M. Birattari, G. Bontempi, and H. Bersini, "Lazy learning meets the recursive least squares algorithm," in *Advances in Neural Information Processing Systems*, M. S. Kearns, S. A. Solla, and D. A. Cohn, Eds. Cambridge, MA: MIT Press, 1999, vol. 11, pp. 375–381.
- [24] G. Bontempi, "Local learning techniques for modeling, prediction and control," Ph.D. dissertation, IRIDIA, Université Libre de Bruxelles, Brussels, Belgium, 1999.
- [25] G. Bontempi, H. Bersini, and M. Birattari, "The local paradigm for modeling and control: From neuro-fuzzy to lazy learning," *Fuzzy Sets Syst.*, vol. 121, no. 1, pp. 59–72, 2001.
- [26] G. Bontempi, M. Birattari, and H. Bersini, "Local learning for iterated time-series prediction," in *Machine Learning: Proceedings of the Sixteenth International Conference*, I. Bratko and S. Dzeroski, Eds. San Francisco, CA: Morgan Kaufmann, 1999, pp. 32–38.
- [27] —, "Lazy learners at work: The lazy learning toolbox," in *Proc. EUFIT'99, 7th Eur. Congr. Intelligent Techniques and Soft Computing*, Aachen, Germany, 1999, CD-ROM.
- [28] M. Birattari and G. Bontempi. (1999) The Lazy Learning Toolbox, For use with Matlab. IRIDIA-Université Libre de Bruxelles. [Online]. Available: <http://iridia.ulb.ac.be/~lazyl/>



**Domenico Villacci** (M'01) was born in Montesarchio, Italy, in 1958. He received the M.Sc. degree in electrical engineering from the University of Naples "Federico II," Naples, Italy, in 1985.

Since 2000, he has been a Full Professor of Electric Power Systems in the Department of Engineering, University of Sannio, Benevento, Italy. His special fields of interest include electric power system analysis with particular emphasis on innovative architectures for remote control and management of electricity distribution systems, protection and

diagnostics of complex systems, integration of distributed generation systems on electrical networks, and grid impacts of renewable sources.



**Gianluca Bontempi** graduated with honors in electronic engineering from the Politecnico di Milano, Milan, Italy, and received the Ph.D. degree in applied sciences from the Universite Libre de Bruxelles (ULB), Brussels, Belgium.

Since January 2002, he has been an Associate Professor in the Computer Science Department at ULB. He has taken part in research projects in academia and in private industries all over Europe. His interests cover data mining, machine learning, bioinformatics and numerical simulation. He has authored more than

30 scientific publications.

Dr. Bontempi is the coauthor of software for data mining and prediction which received awards in two international competitions.



**Alfredo Vaccaro** (M'01) was born in Vico Equense, Italy, in 1974. He received the M.Sc. degree in electronic engineering from the University of Salerno, Salerno, Italy, in 1998.

From 1999 to 2002, he was an Assistant Researcher in the Department of Electrical and Electronic Engineering, University of Salerno. Since March 2002, he has been an Assistant Professor of Electric Power Systems in the Department of Engineering, University of Sannio, Benevento, Italy. His special fields of interest include soft computing and

interval-based methodologies applied to power system analysis and advanced control architectures for diagnostics and protection of distribution networks.



**Mauro Birattari** received the Masters degree in electronic engineering from the Politecnico di Milano, Milan, Italy, in 1997. He is currently working toward the Ph.D. degree at the Universite Libre de Bruxelles, Brussels, Belgium.

He is an active researcher in the field of machine learning. He has coauthored several scientific articles on local learning approaches and on their application to a variety of problems including time-series prediction, system identification, and automatic control.

More recently, he has focused on the application of machine learning techniques to combinatorial optimization and in particular to automatic tuning metaheuristics.