Grey-box models for steam soil disinfestation simulation

R. Berruto, P. Gay∗, P. Piccarolo, C. Tortia
Dipartimento di Economia e Ingegneria Agraria Forestale e Ambientale,
Università degli Studi di Torino via Leonardo da Vinci 44, Grugliasco (TO), Italy

Abstract
Agricultural and biological processes often involve complex phenomena that depend not only on time but also on other independent variables, like spatial coordinates. These processes are intrinsically distributed parameter systems whose modelling requires the adoption of (non linear) partial differential equation. Due to their complexity these models often involve parameters difficult to be estimated, require time-consuming computations and could be not suitable for control purposes.

In this paper, a new model structure for the simulation of steam soil disinfestation processes has been developed. The proposed model is based on a grey-box structure and it is mainly constituted by a couple of lumped linear parameter varying (LPV) switching model.

Attention is focused on the simulation of sheet steaming soil disinfestation processes that are nowadays gaining growing interest as an ecological alternative to methyl bromide fumigation. To this extent, the model has been identified and validated using real data collected during greenhouse and open field steam treatments.

© 2003 IMACS. Published by Elsevier B.V. All rights reserved.

Keywords: Simulation of agricultural processes; Distributed parameter systems; Linear parameter varying models; Steam disinfestation

1. Introduction
In agricultural and biological science complex systems often need to be described with mathematical models. Different approaches for the development of model structure adequate for practical use can be found in literature (see, e.g. [1–3]).

Many environmental and agricultural processes involve mass or energy transport phenomena which behave as intrinsically distributed parameter systems. Their description should be afforded adopting partial differential equations (PDE) that are function of time as well as of spatial coordinates. Finite difference and finite elements techniques allow the approximation of such models by equivalent finite states models driven by ordinary differential equations in which the infinite number of states is replaced

∗ Corresponding author. Tel.: +39.011-6708620; fax: +39.011-6708591.
E-mail address: paolo.gay@unito.it (P. Gay).

0378-4754/$30.00 © 2003 IMACS. Published by Elsevier B.V. All rights reserved.
by a large enough number of states (see, e.g. [4,5]). The resulting models are therefore suitable for scientific and time consuming simulations but are far too complex for on-line applications and control needs.

Distributed parameter systems, considered as systems with a very large number of states could be approximated with low order time-invariant models, as usual in control engineering practice. Simple physical observations on the actual system could help the designer in the choice of the structure of the approximated models. In many cases, the simplified model should maintain some information about the spatial structure of the system, making it reliable also when used with input–output configurations different from the nominal ones.

In this paper, a model structure for the process of soil disinfection by means of steam has been studied. Different methods have been proposed to apply steam to the soil [6]. In this paper, the attention is focused on sheet steaming, which is carried on covering the soil with a thermo-resistant plastic sheet being sealed at the edges [7] under which superheated steam is blown and left to penetrate the soil. This method can be applied in greenhouses as well as in open field. The steam penetration depends on several factors, of which the soil type, the soil conditions and the soil cultivation are the most important ones. For an efficient soil disinfection by steam a temperature of about 70 °C should be maintained for at least 30 min [6,8] in all the soil layers interested by the roots of the cultivated species. The costs are actually a hard constraint to the feasibility of sheet steaming and primarily consist in the expense of the fuel needed to generate the steam.

The study of the thermal behavior of the soil during the treatment is hard to afford with standard techniques. The process results to be an intrinsically nonlinear distributed parameter system where the temperature depends both on the variables time and depth.

The model structure proposed in this paper follows a fairly new approach. Instead of relying on physical-based models (potentially highly complex and nonlinear), is based on a couple of lumped parameter single-input single-output linear parameter varying (LPV) (see, e.g. [9–12]) dynamic models that directly relate the input signal of the steam valve command with the soil temperature at different depths. The limited complexity and the compactness of the proposed model structure makes it suitable for on-line application and control purposes [13].

In order to identify the proposed model, input–output data (the signal of the valve and the temperature measurements at different depth in the soil) have been collected from test bench in greenhouse and open field trials.

The paper is structured as follows: Section 2 reports a brief overview about physical phenomena involved during the steam disinfection process. In Section 3, the grey-box model structure is introduced and discussed, while in Section 4 are reported some topics about the identification of the blocks constituting the model, together with some details about the performed experiments and achieved results. Section 5 concludes the paper.

2. The steam disinfection: description of the process

Sheet steaming can be performed either in greenhouse or in open field. In both the cases the soil is covered with a sheet anchored to the soil or to the greenhouse bench and steam is injected by means of cloth hoses. In Fig. 1, a diagram of the treatment in a greenhouse bench and open field, respectively, are reported.
The steam disinfection process can be separated in two successive phases, in which the involved physical phenomena are different. During the first one, referred to as the heating phase, the steam produced in the boiler is supplied to the cloth hoses and then flows through the soil. The second phase, referred to as cooling, concerns the free evolution of the system following the heating process.

During the heating process, the usual dynamics that regulate the heat transport (thermal conduction, convection and radiation) are combined with water vapor diffusion phenomena. The steam is injected under the heat-resistant film by means of one or more cloth hose longitudinally perforated and progressively expands, inflating the film. Steam loss are very little so that almost all the steam flows through the soil or condensates. The pressure achieved under the sheet mainly depends on the steam flow rate supplied by the pipe, the partial condensation of the steam, the air and soil temperature, the steam flux through the soil, the weight of the sheet and the resistance due to its deformation. Once the steady state operating condition is achieved, the sheet is fully inflated and the resulting pressure under the sheet makes the diffusion of the steam through the soil quite homogeneous. Neglecting boundary effects is possible to assume the temperature of the soil as dependent, from a geometrical point of view, on the depth $\xi$ only.

During the cooling phase, water vapor diffusion, which is still present at the time of closure of the valve, progressively diminishes. The absence of the steam flow during the cooling results in slower dynamics than in the heating phase. An example of the behavior of the soil temperatures, measured during a treatment, can be found in Fig. 2. The different dynamics involved in the heating and cooling phase evidence a structural non-linearity of the process. Both heating and cooling dynamics depend on the depth. Generally heating/cooling rate is faster at the surface and slows down as the depth increases.

Unfortunately the physical modeling of the whole process concerns various interdependent conditions and combined modes (see, e.g. [14,15]) that necessarily lead to heavily involved nonlinear partial differential equations, that could be too complex for on line applications. Moreover, due to high number of physical parameters involved, this model could result to be critical to be identified [16].

The model, once identified, is suitable for the simulation of soil temperature at the particular conditions, among which soil composition, texture and moisture content, observed when collecting measurements for the identification. However, this is not a limitation because, in the matter of fact, farmers perform the treatment periodically on the same type of soil and tillage and when the water content is fairly low to have good steam permeability of the soil. This fact reduces variability in the treatments, improving the reliability of the model.
3. The switching multiple LPV based model

The objective of this study was to develop a reliable model, whose structure is simple enough to allow the on-line prediction of the soil temperature, which is an essential task for control purpose [13].

The input of the model is the binary steam valve opening control signal \( u(k) \), while the output \( y(k; \xi) \) is the temperature of soil at a given depth \( \xi \) and time instant \( k \). A preliminary experimental study confirmed the homogeneous behavior of the temperature of soil in different positions in the plot, which definitely results to depend, from a geometrical point of view, on one coordinate only (the depth).

Due to the different dynamics involved in the heating and cooling phases, at least two different (linear) models should be considered to describe the behavior of the soil temperature during the complete treatment. Each of these subsystems is designed to take care of one of the above defined phases.

To this aim, the two subsystems-based switching model in Fig. 3 has been adopted. Each subsystem has been modelled by an LPV model which consists of a linear lumped parameter model in which the parameters are not constant, but are functions of an extra, possibly vector valued, variable that can be regarded as an input determining the “operating condition” of the model (see, e.g. [9,10,17,18]). In our settlement, the depth \( \xi \) has been assumed to be the external parameter of the two LPV models. This modelling approach allows to describe the forced system response of intrinsically distributed parameter systems without involving partial differential equations.

The switching command signal has been assumed to be driven by the closing of the steam valve. Actually, input–output observations yield the presence of a heat transfer delay depending on the depth \( \xi \). Such depth dependent delay, that should be separately taken into account when dealing with linear models,
affects the input $u$ and therefore the switching between the heating and cooling models. To this extent, the delayed signal $u_{\Delta}(k)$, driving both the subsystems and the switch, has been introduced according to

$$u_{\Delta}(k) = u(k - \Delta(\xi)) = q^{-\Delta(\xi)}u(k),$$

where $\Delta(\xi)$ is an unknown continuous function to be estimated and $q^{-1}$ is the unit time delay operator.

With this assumption, the switching between the two models for soil layers at different depths generally does not occur at the same instant.

More rigorously, the involved two subsystems can be expressed as follows.

The LPV heating subsystem has an auto-regressive (ARX) structure with a moving average exogenous input

$$A_h(q^{-1}, \xi)y_h(k; \xi) = B_h(q^{-1}, \xi)u_{\Delta}(k) + e(k),$$

where

$$A_h(q^{-1}, \xi) = 1 + a_{h1}(\xi)q^{-1} + a_{h2}(\xi)q^{-2} + \cdots + a_{hna}(\xi)q^{-na},$$

$$B_h(q^{-1}, \xi) = b_{h0}(\xi) + b_{h1}(\xi)q^{-1} + b_{h2}(\xi)q^{-2} + \cdots + b_{hnb}(\xi)q^{-nb}.$$  

What makes the difference between LPV and standard ARX models is the fact that parameters $a_{hi}(\xi), i = 1, \ldots, na_h$, and $b_{hi}(\xi), i = 0, \ldots, nb_h$, are not constant and are unknown (continuous) functions of the depth $\xi$ to be estimated.

The structure of the cooling subsystem is the same, that is

$$A_c(q^{-1}, \xi)y_c(k; \xi) = B_c(q^{-1}, \xi)u_{\Delta}(k) + e(k),$$

with

$$A_c(q^{-1}, \xi) = 1 + a_{c1}(\xi)q^{-1} + a_{c2}(\xi)q^{-2} + \cdots + a_{nc}(\xi)q^{-nc},$$

$$B_c(q^{-1}, \xi) = b_{c0}(\xi) + b_{c1}(\xi)q^{-1} + b_{c2}(\xi)q^{-2} + \cdots + b_{nc}(\xi)q^{-nb},$$

where, again, the functions $a_{ci}(\xi), i = 1, \ldots, na_c$, and $b_{ci}(\xi), i = 0, \ldots, nb_c$, are to be estimated.
The switching subsystem is driven by the binary signal \( u_{\Delta_1}(k) \) and generates the global output \( y(k; \xi) \) according to the following equation:

\[
y(k; \xi) = \begin{cases} 
  y_h(k; \xi), & \text{if } u_{\Delta_1}(k) = 1, \\
  y_c(k; \xi), & \text{if } u_{\Delta_1}(k) = 0.
\end{cases}
\] (6)

### 4. Model identification

The model has been identified and validated using real data collected from a set of experiments both in greenhouse and in open field. Soil disinfections were performed in usual operating conditions as reported in Table 1.

The temperature of the soil was measured using probes equipped by seven thermocouple sensors placed at different depths (15, 40, 65, 90, 115, 140 and 165 mm). The probes have been connected to a data-logger collecting measures with a sampling rate of 10 s. The performances of the steam generator are reported in Table 2.

During the treatment, the partial condensation of the steam increases the soil moisture. This changes the thermal and physical properties of the soil, making the system non-stationary. A reliable model could be obtained identifying the system at the same operating conditions of the actual disinfection operation, including the command signal \( u \). This fact introduce severe constraints on the design of the input signal \( u \). To this extent we adopted the following input signal:

\[
u(k) = \begin{cases} 
  0, & k \leq 0, \\
  1, & 1 \leq k \leq k_{stop}, \\
  0, & k > k_{stop}.
\end{cases}
\] (7)

<table>
<thead>
<tr>
<th>Treatment data sheet</th>
<th>Greenhouse</th>
<th>Open field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface treated (m²)</td>
<td>40</td>
<td>400</td>
</tr>
<tr>
<td>Initial water content (%)</td>
<td>13</td>
<td>9.7</td>
</tr>
<tr>
<td>Treatment duration (h)</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Cloth hoses number</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Type of soil</td>
<td>Medium</td>
<td>Medium, well drained</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>Boilers data-sheet</th>
<th>Greenhouse</th>
<th>Open field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steam production (kg h⁻¹)</td>
<td>200</td>
<td>2000</td>
</tr>
<tr>
<td>Steam temperature (°C)</td>
<td>170</td>
<td>170</td>
</tr>
<tr>
<td>Gas oil consumption (kg h⁻¹)</td>
<td>13</td>
<td>167</td>
</tr>
</tbody>
</table>
which has the same structure of the usual valve command adopted in manual operations. Unfortunately, input signals of this kind turn out to be not very rich from a spectral point of view, and in most cases the length of the opening valve phase, \( k_{\text{stop}} \), steps, could not allow the system to reach a steady state condition at all depths.

The experimental data were first divided in two subsets. The first one, used for the identification of the model, consisted in the temperature measurements at the depths 15, 40, 65, 115, 140 and 165 mm, while the measurements at 90 mm were left out for validation purposes. The first step in the identification process concerned the determination, by means of standard techniques, of the input delay at each depth. The function \( \Delta(\xi) \) was then obtained interpolating the six estimated delays with a cubic spline. An example of the estimated function \( \Delta(\xi) \) is reported in Fig. 4.

The two LPV subsystems can then be separately identified. To this extent we proceeded to a data allocation step, consisting in separating the identification data into two groups, one for each subsystem. The data used for the identification of the heating subsystem consisted in the data collected from the valve opening time to the valve closing time, shifted according to the corresponding estimated delay \( \Delta(\xi) \). The data for the cooling subsystem identification were the remaining ones.

The error in relations (2) and (4) has been assumed to be unknown but bounded, that is, if we consider a data set consisting of \( m \) measurements, the error vector \( e \in \mathbb{R}^m \) necessarily belongs to a membership set

\[
\Omega_e = \{ e \in \mathbb{R}^m : |e(k)| \leq E(k), \forall k \}
\]

where \( E(k) \) is a user-defined bounding error function. For any subsystem, six different models have been identified. The LPV systems, when considered for fixed \( \xi \), results to be ARX models. For what concerns the heating subsystem, at any depth \( \xi_j, j = 1, \ldots, 6 \), the following ARX model, derived from (2) has
been identified

\[ y_h(k; \xi) = -\sum_{i=1}^{na_h} a_h(\xi) y_h(k-i) + \sum_{j=0}^{nb_h} b_h(\xi) u_\Delta_1(k-j). \]

Parameters \( a_h(\xi), i = 1, \ldots, na_h \), and \( b_h(\xi), j = 0, \ldots, nb_h \), can be arranged, for notation convenience, into a single vector \( \theta_h(\xi) \in \mathbb{R}^{np_h} \),

\[ \dot{\theta}_h(\xi) = \left[ \begin{array}{c} a_h(\xi) \\ \vdots \\ a_{na_h}(\xi) \\ b_h(\xi) \\ \vdots \\ b_{nb_h}(\xi) \end{array} \right]^T \]

so that we can write

\[ y_h(\xi) = \Phi_h \theta_h(\xi) + e, \]

where \( y_h(\xi) \in \mathbb{R}^m \) and \( e \in \mathbb{R}^m \) are the measurement and the equation error vectors, respectively, and \( \Phi_h \) is the regression matrix, whose \( i \)th row \( \phi_i^T \) is given by

\[ \phi_i^T = [-y_h(i-1; \xi), \ldots, y_h(i-na_h; \xi), u_\Delta_1(i) \cdots u_\Delta_1(i-nb_h)], \quad i = 1, \ldots, m. \]

With the set membership error assumption of relation (8), the identification of the parameter vector \( \theta_h(\xi) \) consists in finding the set \( D_{\theta_h}(\xi) \) of all parameter vectors \( \theta_h \) consistent with the model (10), the measurements \( y_h \) and the errors \( e \).

The corresponding parameter admissible set \( D_{\theta_h}(\xi) \) can be expressed as

\[ D_{\theta_h}(\xi) = \{ \theta_h \in \mathbb{R}^{np_h}, y_h = \Phi_h \theta_h(\xi) + e, e \in \Omega_e \}. \]

Any point in \( D_{\theta_h}(\xi) \) can, in principle, be an estimate of the parameter vector while the “size” of \( D_{\theta_h}(\xi) \) is a measure of the parameter reliability. From relations (8) and (12), it follows that the parameter admissible set \( D_{\theta_h}(\xi) \) is a polytope described by a subset of the following \( 2m \) inequalities

\[ y_h(k; \xi) - E(k) \leq \phi_i^T \theta_h(\xi) \leq y_h(k; \xi) + E(k), \quad i = 1, \ldots, m. \]

Several point estimate, satisfying different optimality criterion, have been studied in literature (see, e.g. [19]). In this paper, the \( \ell_\infty \) optimal projection estimate \( \hat{\theta}_h \in D_{\theta_h}(\xi) \) defined as

\[ \hat{\theta}_h = \arg \min_{\theta_h \in D_{\theta_h}(\xi)} || \Phi_h \theta_h - y_h ||_\infty, \]

has been adopted. Such an estimate can be computed solving a linear programming problem involving \( np_h + 1 \) variables.

The initial conditions of the heating model were chosen to correspond to the initial soil temperature \( T_0, i.e. \)

\[ y_h(-na_h+1; \xi) = \cdots = y_h(0; \xi) = T_0 = 20^\circ C, \quad \forall \xi. \]

A similar formalism was adopted for the identification procedure of the cooling LPV system. The main difference relies in the fact that, due to the switching action, the initial conditions of the cooling model

\[ \text{The } \ell_\infty \text{ norm is defined as } ||y||_\infty = \max_i |y_i|. \]
should correspond to the output temperatures of the heating process, i.e. if we denote by $k^*(\xi_j)$ the switching instant at depth $\xi_j$, it results

$$[y_c(k^*(\xi_j) - n_a^c + 1) \cdots y_c(k^*(\xi_j))] = [y_h(k^*(\xi_j) - n_a^h + 1) \cdots y_h(k^*(\xi_j))]$$

The consequences of the weak excitation properties and the short duration of the input signal could be overcome by a suitable setting of the weighting function $v$ in relation (14) that, from a practical point of view, could be performed assuming a proper bounding function $E(k)$.

Different model orders were considered, evaluating both the one-step prediction error and, more importantly, the simulation error. The adopted orders for the final model were $n_a^h = 4$, $n_b^h = 1$ for the heating system and $n_a^c = 4$, $n_b^c = 1$ for the cooling one. These model orders were the lowest at which simulation error was considered acceptable at all the measured depths.

The $a_i(\xi)\ i = 1, \ldots, n_a^h$, and $b_i(\xi)\ i = 0, \ldots, n_b^h$ functions for the heating subsystem, and the corresponding ones for the cooling subsystem, were then obtained interpolating with cubic splines the coefficients $a_i(\xi_j)\ i = 1, \ldots, n_a^c$, and $b_i(\xi_j)\ i = 0, \ldots, n_b^c$, for $j = 1, \ldots, 6$, obtained identifying models in relations (2) and (4) at the six measuring depths.

To validate the model, the obtained LPV models were then used to predict the temperature behavior for the depth of 90 mm. The simulated results, for the cooling phase, are presented in Fig. 5. As can be seen in greenhouse as in open field cases the achieved simulation results are precise enough for practical application like on-line control or simulation for process parameter optimization.

5. Conclusions and perspectives

In this paper an input–output grey-box model for the simulation of the temperature of soil during steam disinfection processes has been presented. The structure of the model has been designed on the base of empirical considerations on the actual system and it is based on the interconnection of a number of
linear blocks. The spatial coordinates dependence is taken into account by two linear parameter varying models which commute according to the general state of the system.

The proposed modelling approach allows to describe forced system response of intrinsically distributed parameter systems without involving partial differential equations. The resulting model is therefore easy to be identified and requires small computational resources for its simulation, making it viable for on-line application, like prediction and control.

The reported results about the identification and the validation of the resulting model on the basis of real experimental data collected both from a greenhouse test bench and from open field show how the proposed model structure is suitable for on-line applications where the prediction and the simulation of the forced response to lumped inputs is of interest.

Acknowledgements

This work was partially supported by the grants of the Assessorato per l’ Agricoltura of the Regione Piemonte, Italy.

References