

# Temperature Prediction in District Heating Systems with cFIR models

**Pierre Pinson\***, Torben S. Nielsen, Henrik Aa. Nielsen, Niels K. Poulsen & Henrik Madsen  
Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark

## Abstract

Current methodologies for the optimal operation of district heating systems are based on model predictive control. In complement to load forecasts, accurate predictions (up to 12-hour ahead) of the water temperature at critical points of the networks are crucial for meeting constraints related to consumers while minimizing the production costs for the heat supplier. The paper introduces a new forecasting methodology based on a conditional Finite Impulse Response (cFIR) model, for which the model coefficients are replaced by nonparametric or semi-parametric coefficient functions of the water flux at the supply point and of the time of day. This allows for nonlinear variations of the time delays in the FIR model. The coefficients functions can be adaptively estimated with a method that combines local polynomial regression, exponential forgetting, recursive weighted least squares and Tikhonov regularization. Results are given for the test case of the Roskilde district heating system, over a period of more than 6 years. The advantages of the proposed forecasting methodology in terms of a higher forecast accuracy, in terms of its use for simulation purposes, or alternatively for better understanding transfer functions of district heating systems, are clearly shown.

**Key words:** district heating, control, forecasting, time delay, finite impulse response, coefficient functions, adaptive estimation

\* *Corresponding author:*

P. Pinson, [Informatics and Mathematical Modelling, Technical University of Denmark](#),  
Richard Petersens Plads (bg. 321 - 020), DK-2900 Kgs. Lyngby, Denmark.  
Tel: +45 4525 3428, fax: +45 4588 2673, email: [pp@imm.dtu.dk](mailto:pp@imm.dtu.dk), webpage: [www.imm.dtu.dk/~pp](http://www.imm.dtu.dk/~pp)

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Description of the proposed forecasting methodology</b>	<b>4</b>
2.1	The state-of-the-art statistical approach . . . . .	4
2.2	The proposed forecasting methodology . . . . .	5
2.2.1	Modeling the transfer function of the network . . . . .	5
2.2.2	Integrating the social behaviour of consumers . . . . .	5
2.2.3	Forecasting with the cFIR models . . . . .	6
<b>3</b>	<b>Estimation of the model parameters</b>	<b>8</b>
3.1	Local polynomial estimates . . . . .	8
3.2	Offline estimation of the local coefficients . . . . .	9
3.3	Online estimation of the local coefficients . . . . .	11
3.4	Regularization for a better generalization ability . . . . .	12
<b>4</b>	<b>Case-studies and results</b>	<b>14</b>
4.1	Description of the case-studies . . . . .	14
4.2	Setup and optimal fitting of the cFIR models . . . . .	14
4.3	Out-of-sample evaluation of the cFIR models . . . . .	18
4.4	Sensitivity to the flux values used as input to the cFIR models . . . . .	19
4.4.1	Flux prediction at the supply point . . . . .	19
4.4.2	Performance of cFIR models fed with flux predictions . . . . .	21
4.5	Comparison with the state-of-the-art approach . . . . .	22
<b>5</b>	<b>Concluding remarks</b>	<b>24</b>
	<b>Acknowledgments</b>	<b>27</b>
	<i>Bibliography</i>	<b>27</b>
<b>A</b>	<b>Detailed evaluation: cFIR models with flux measurements</b>	<b>29</b>
A.1	Critical point 1 . . . . .	29
A.2	Critical point 2 . . . . .	30
A.3	Critical point 3 . . . . .	31
<b>B</b>	<b>Detailed evaluation: cFIR models with flux predictions</b>	<b>32</b>
B.1	Critical point 1 . . . . .	32
B.2	Critical point 2 . . . . .	33
B.3	Critical point 3 . . . . .	34
<b>C</b>	<b>Detailed evaluation: lagged transfer function model</b>	<b>35</b>
C.1	Critical point 1 . . . . .	35
C.2	Critical point 2 . . . . .	36
C.3	Critical point 3 . . . . .	36

# 1 Introduction

District heating systems consist of centralized heat production facilities with associated distribution networks. They play an important role in Nordic countries, where they are used for meeting the demand related to space heating and hot tap water consumption. Owing to this centralized production combined with complex network architectures, decisions made from the supply point of view have highly significant economical impacts. In order to optimally operate district heating systems, control strategies are implemented with some restrictions e.g. a minimum guaranteed inlet temperature at the consumers. The aim of these control strategies is to meet these restrictions while minimizing the supply temperature, and thus the production costs for the heat supplier. Alternative methodologies based on predictive control have been described by [Nielsen \(2002\)](#) and [Sandou et al. \(2004\)](#).

If considering a unique heat supplier, his decision variables are the magnitude of the water flux and the supply temperature. The magnitude of the water flux is directly imposed by the load, and thus load forecasts serve as a basis for making a decision regarding this flux. The accuracy of load forecasts has been discussed by [Nielsen & Madsen \(2000\)](#) or [Dotzauer \(2002\)](#) for instance. Similarly, predictive control for the supply temperature necessitates a model that permits to forecast the water temperature at critical points of the network considered. The relevant forecast horizons may be up to 12-hour ahead. Increasing the accuracy of these temperature forecasts is expected to significantly lower the production costs for the heat suppliers, as a consequence of them making more optimal control decisions. The aim of the present paper is to contribute to reaching a higher forecasting accuracy by proposing a new forecasting methodology.

The models in the literature either derive from a physical description of the heat and mass transfers in the network ([Sandou et al. 2005](#)), or they are based on a statistical description of the transfer function from the supply point to the critical point considered ([Søgaard 1993](#)). In both cases, a fixed time delay in the network is assumed, owing to computational costs or estimation complexity of using a varying time delay. In contrast, the forecasting methodology introduced here permits to account for a varying time-delay in the network. The embedded model is a Finite Impulse Response (FIR) for which the model coefficients are replaced by nonparametric coefficient functions of influential variables. Owing to this consideration of the nonlinear influence of external factors on the FIR, the model is referred to as conditional Finite Impulse Response (cFIR). The proposed cFIR models are used here for capturing the nonlinear influence of the water flux at the supply point on the transfer function of the district heating system. Another interest of cFIR models is that they can account for the influence of the social behavior of the consumers on the temperature at the critical point. It may be done either by having the time of the day as an influential variable of the cFIR model, or more classically, by having an offset term in the form of a diurnal harmonic. Considering these two alternatives will allow us to discuss the assumption of the social behavior of the consumers impacting (or not) the transfer function of the district heating system.

The problem of predicting the water temperature at critical points of a district heating system is described in Section 2, as well as the proposed forecasting methodology. Then, Section 3 introduces the method for the estimation of the model coefficients. Particularly, it allows for a recursive estimation of the coefficients so that it accommodates long-term variations. In addition, a regularization of the recursive estimation method is proposed for enhancing its generalization ability and for controlling its multi-step ahead accuracy. The

case-study of the district heating system of Roskilde in Denmark is considered in Section 4 in order to illustrate the benefits of this new forecasting methodology. The original dataset includes temperature and flux measurements at the supply point, as well as temperature measurements at 3 critical points of the distribution network, for a period of more than 6 years. In addition to demonstrating the significantly higher performance of the proposed methodology, its interest for better understanding the time delays in the network is discussed. Section 5 ends the paper by summarizing the main conclusions and gathering perspectives regarding future developments.

## 2 Description of the proposed forecasting methodology

A district heating system often consists in a complex network. Though, one is mainly interested in what occurs at some specific points of this network. They serve as references for designing and optimizing the control strategies, and are thus referred to as critical points. Focusing on a single critical point, the overall network is conceptually simplified: it is considered that there is a unique simple pipe between the supply and critical points. The district heating system operator injects in a continuous manner quantities of warm water (at a controlled temperature), and is interested in knowing what will be the temperature at this critical point depending on its operation strategy at the supply point. Denote by  $x_t$  and  $u_t$  the value of the flux and of the water temperature at the supply point at time step  $t$ . In parallel,  $y_t$  is the water temperature at the critical point considered at that same time. The problem is here discretized. For practical applications, the suitable time step is typically of one hour. The available data hence consist in the time-series  $\{x_t\}$ ,  $\{u_t\}$  for the supply point and  $\{y_t\}$  for the critical point, all including  $n$  observations. The state-of-the-art statistical approach is introduced in a first part, followed by the description of our new forecasting methodology based on a cFIR model

### 2.1 The state-of-the-art statistical approach

The model that is traditionally used for predicting the temperature at critical points, initially proposed by Sogaard (1993), takes the form of a linear transfer function model with a first-order autoregressive component

$$y_t = a_1 y_{t-1} + b_0(h_t - \tau)u_{t-\tau} + b_1(h_t - \tau - 1)u_{t-\tau-1} + b_2(h_t - \tau - 2)u_{t-\tau-2} + \varepsilon_t, \quad \forall t \quad (1)$$

where  $h_t$  is the hour of the day corresponding to the time step  $t$ ,  $\{\varepsilon_t\}$  is the noise sequence, such that  $\mathbb{E}[\varepsilon] = 0$  and  $\sigma_\varepsilon^2 < \infty$ . While  $a_1$  is not conditional to any variable, the coefficients  $b_j$ ,  $j = 0, 1, 2$ , of the transfer function are made a function of the time of the day and of the time delay  $\tau$  in the system, by using a Fourier harmonics with a period of 24 hours

$$b_j(h_t - \tau - j) = b_j^0 + b_j^1 \sin\left(\frac{\pi(h_t - \tau - j)}{12}\right) + b_j^2 \cos\left(\frac{\pi(h_t - \tau - j)}{12}\right), \quad \forall t, \quad j = 0, 1, 2 \quad (2)$$

i.e. they account for the diurnal variations in the system behavior, owing to the social behavior of the consumers. When using this model for predictive control, Nielsen (2002) proposed to choose  $\tau$  as the time delay that maximizes the correlation between the time series  $\{y_t\}$  and  $\{u_{t-\tau}\}$ . It can be allowed to change over time by using a sliding window, thus yielding different time delays in the system depending on the season of the year.

The above model may provide an acceptable description of  $\{y_t\}$ . Though, if used for multi-step ahead prediction, the forecast accuracy lowers dramatically as the lead time gets further. This is due to the autoregressive part of the model. Indeed, when issuing at time  $t$  a  $k$ -step ahead forecast  $\hat{y}_{t+k|t}$ , the model is fed with the  $(k-1)$ -step ahead forecast  $\hat{y}_{t+k-1|t}$ . A consequence is that forecasting errors directly sum up as  $k$  increases. This will be illustrated in Section 4, where model (1) will be used as a benchmark. Another drawback of the model, which significantly affects its performance when used for forecasting purposes, is the fixed time delay  $\tau$ , which is not realistic. In practice,  $\tau$  does not only vary depending on the season: it is necessarily a function of the flow in the network (Arvatson 2001).

## 2.2 The proposed forecasting methodology

### 2.2.1 Modeling the transfer function of the network

Owing to the drawbacks of model (1), it is proposed to introduce a new model with a varying time-delay, and without any autoregressive component. For modeling the transfer function of the network, we use as a basis the conditional Finite Impulse Response (cFIR) model initially described by Nielsen (2000), i.e.

$$y_t = \sum_{j \in S_j} \beta_j(x_{t-1}) u_{t-j} + \varepsilon_t, \quad \forall t \quad (3)$$

where  $y_t$  is the temperature at the considered critical point at time  $t$ ,  $x_{t-1}$  is the value of the flux at the supply point at time  $t-1$  and  $u_{t-j}$  are the lagged values of the temperature at the supply point.  $S_j$  corresponds to the finite set of indexes related to the lagged values for the cFIR model.  $\{\varepsilon_t\}$  is a white noise sequence, for which  $\mathbb{E}[\varepsilon_t] = 0$  and  $\sigma_{\varepsilon_t}^2 < \infty$ .

The advantage of model (3) is that the cFIR is conditional to the flux at the supply point, since the coefficients  $\beta_j$  are indeed coefficient functions of  $x_{t-1}$ . This way, the time delay in the system is also made a (nonlinear) function of the flux. However, since the model does not integrate any component describing the autocorrelation of the  $\{y_t\}$  time-series, it may require the cardinal of  $S_j$  to be large. In addition, it does not account for potential heat losses that would be a function of the magnitude of the flux. For these two reasons, it is proposed here to add an offset term (also function of the flux) in model (3), so that it becomes

$$y_t = \beta_0(x_{t-1}) + \sum_{j \in S_j} \beta_j(x_{t-1}) u_{t-j} + \varepsilon_t, \quad \forall t \quad (4)$$

### 2.2.2 Integrating the social behaviour of consumers

Model (4) is expected to provide an adequate description of the flux-dependent transfer function of the distribution network. However, it does not account for the potential influence of the social behavior of the consumers, i.e. for their consumption pattern (as a function of the hour of the day) that necessarily depends on the type of consumers that are connected to the critical point considered. It is intuitive that the consumption pattern is very different if the critical point corresponds to a residential area, to an industrial area, or to an hospital.

There may be two alternative views on how to integrate the social behaviour of the consumers within model (4). On the one hand, one may consider that it does not impact the transfer function of the network. In this case, it is only necessary to modify the offset term of the model, in order for it to account for diurnal variations. It is proposed here to use Fourier harmonics of period 24 hours, so that the offset term  $\beta_0$  in (4) is replaced by

$$\beta_0^\dagger(x_{t-1}, h_t) = \beta_{0,0}^\dagger(x_{t-1}) + \beta_{0,1}^\dagger(x_{t-1}) \sin\left(\frac{\pi h_t}{12}\right) + \beta_{0,2}^\dagger(x_{t-1}) \cos\left(\frac{\pi h_t}{12}\right), \quad \forall t \quad (5)$$

with  $h_t$  the hour of the day at time step  $t$ , while the other  $\beta_j^\dagger$  coefficient functions remain unchanged

$$\beta_j^\dagger(x_{t-1}) = \beta_j(x_{t-1}), \quad \forall j, j > 0 \quad (6)$$

This model will be referred to as a ‘rigid’ cFIR in the following.

The  $\beta_j^\dagger$  coefficient functions can be gathered in vector denoted by  $\beta^\dagger(x_{t-1}, h_t)$ . In parallel, let  $\mathbf{u}_t^\dagger$  be the corresponding vector of ones, harmonics values, as well as lagged values of  $u_t$ . Then, the rigid cFIR model can simplify to

$$y_t = \beta^{\dagger\top}(x_{t-1}, h_t) \mathbf{u}_t^\dagger + \varepsilon_t^\dagger, \quad \forall t \quad (7)$$

On the other hand, one may consider that the social behavior of the consumer also influences the transfer function of the network. Then, this translates to having the  $\beta_j$  coefficients functions in (4) being a function of the time of the day, in addition to being a function of the flux at the supply point. This yields the alternative model

$$y_t = \beta_0^*(x_{t-1}, h_t) + \sum_{j \in S_j} \beta_j^*(x_{t-1}, h_t) u_{t-j} + \varepsilon_t^*, \quad \forall t \quad (8)$$

that is, with the  $\beta_j^*$  coefficient functions being a function of both the flux  $x_{t-1}$  at the supply point and the time of the day  $h_t$ . Owing to its more supple structure, the model will be referred to as a ‘supple’ cFIR in the following. In the same manner than for the rigid one, denote by  $\beta^*(x_{t-1}, h_t)$  the vector of coefficient functions for this model and by  $\mathbf{u}_t^*$  the corresponding vector of ones and lagged values of  $u_t$ . In such case, model (8) becomes

$$y_t = \beta^{*\top}(x_{t-1}, h_t) \mathbf{u}_t^* + \varepsilon_t^*, \quad \forall t \quad (9)$$

To sum up, the difference between the supple and rigid cFIR models is that the former ones have their coefficient functions conditional to both the flux at the supply point and the time of the day, while for the latter ones they are only conditional to flux values.

### 2.2.3 Forecasting with the cFIR models

Models (7) and (9) describe the temporal evolution of  $\{y_t\}$  from past information at the supply point, i.e. measurements of flux and supply temperature. From these models, the one-step ahead prediction at time  $t$  of the temperature at the critical point can be defined as the conditional expectation of the process at time  $t + 1$  given the information set  $\Omega_t$  up to time  $t$ , and the chosen cFIR model. If denoting by  $\hat{y}_{t+1|t}^\dagger$  the one-step ahead prediction

for the rigid cFIR model, this writes

$$\hat{y}_{t+1|t}^\dagger = \mathbb{E} [y_{t+1} | \beta^\dagger, \Omega_t], \quad \forall t \quad (10)$$

or alternatively

$$\hat{y}_{t+1|t}^* = \mathbb{E} [y_{t+1} | \beta^*, \Omega_t], \quad \forall t \quad (11)$$

for the supple cFIR model, with  $\hat{y}_{t+1|t}^*$  denoting the corresponding one-step ahead prediction.

In practice, since  $\mathbb{E} [\varepsilon^\dagger] = 0$  and  $\mathbb{E} [\varepsilon^*] = 0$ , the one-step ahead predictions are obtained with

$$\hat{y}_{t+1|t}^\dagger = \beta^\dagger(x_t, h_{t+1})^\top \mathbf{u}_t^\dagger, \quad \forall t \quad (12)$$

and

$$\hat{y}_{t+1|t}^* = \beta^*(x_t, h_{t+1})^\top \mathbf{u}_t^*, \quad \forall t \quad (13)$$

for the rigid and supple cFIR models, respectively.

Note that for computing one-step ahead predictions with cFIR models, the current filtered flux value  $x_t$  and the lagged supply temperature values  $u_{t-j+1}$ ,  $j \in S_j$ , are available measurements. But then, for multi-step ahead prediction, say  $k$ -step ahead, the above equations become

$$\hat{y}_{t+k|t}^\dagger = \beta^\dagger(\hat{x}_{t+k-1|t}, h_{t+k})^\top \mathbf{u}_{t+k-1}^\dagger, \quad \forall t \quad (14)$$

and

$$\hat{y}_{t+k|t}^* = \beta^*(\hat{x}_{t+k-1|t}, h_{t+k})^\top \mathbf{u}_{t+k-1}^*, \quad \forall t \quad (15)$$

This means that since  $u$  is the control variable, one can assume that future values are known. In contrast, future flux values at the supply point cannot be exactly known, though they could be deduced from load predictions, since the load  $q_t$  at time  $t$  is related to  $x_t$  with

$$q_t = c_p x_t (u_t - v_t), \quad \forall t \quad (16)$$

where  $v_t$  is the return temperature at that same time and  $c_p$  is the heat capacity of the water in the pipe. It is known that the variations of  $v_t$  are very smooth, so that future values can be assumed to be known or could be accurately predicted with e.g. exponential smoothing. Consequently, predictions of future flux at the supply point could be straightforwardly obtained and used to feed the model.

Actually, both flux and temperature at the supply point are control variables in practice, which are interdependent. For instance, for reaching the same temperature at the critical point, increasing the flux may allow to have a lower supply point temperature. Restrictions on the range of potential values and variations of these two variables may also step in this complex decision-making problem. Anyways, the cFIR models can also be used for simulation purposes in order to evaluate the impact of decisions on the flux and temperature variables in the following hours. In order to illustrate the sensitivity of the prediction performance of the cFIR models to the choice of future flux values, we will consider in Section 4 a forecasting exercise where the supply temperature is the unique control variable



and for which a simple Auto-Regressive (AR) model of order  $p$  is used for describing  $\{x_t\}$ , and consequently for multi-step ahead prediction

$$x_t = \alpha_0 + \sum_{i=1}^p \alpha_i x_{t-i} + \xi_t, \quad \forall t \quad (17)$$

where  $\{\xi_t\}$  is a white noise sequence, i.e. such that  $\mathbb{E}[\xi] = 0$  and  $\sigma_\xi^2 < \infty$ . Such modeling approach may be less appropriate than that mentioned above, so that the results given in Section 4 will consist a lower bound on the potential performance of the proposed forecasting methodology in comparison to the case for which more advanced flux predictions would be available. This will permit to perform a sensitivity analysis on the prediction performance of the cFIR models even if the information on future flux at the supply point is not perfectly accurate. After inspection of the correlogram of the model residuals, it has been decided to enhance this AR( $p$ ) model with a Fourier harmonics of period 24 hours. This allows us to account for a periodic diurnal variations in the  $\{x_t\}$  time series that cannot be captured by the autoregressive component only. Model (17) becomes

$$x_t = \alpha_0^0 + \alpha_0^1 \sin\left(\frac{\pi h_t}{12}\right) + \alpha_0^2 \cos\left(\frac{\pi h_t}{12}\right) + \sum_{i=1}^p \alpha_i x_{t-i} + \xi_t, \quad \forall t \quad (18)$$

where  $h_t$  is the hour of the day corresponding to time step  $t$ . Multi-step ahead flux forecasts obtained from model (18) are then used to feed the cFIR model as expressed in (14) and (15).

### 3 Estimation of the model parameters

For a model such as (18), the model parameters can be easily estimated with a Least Squares (LS) or Recursive Least Squares (RLS) method, as described in e.g. (Madsen 2006). In contrast, for the case of cFIR models, the chosen method for nonparametric parameter estimation is described in the following, both for offline and online applications. It combines local polynomial regression, weighted LS for the offline case – and respectively weighted RLS with exponential forgetting for the online case, as well as Tikhonov regularization. For simplicity, the method is described for a generic cFIR model whose transfer function is described by the  $\beta(\mathbf{r})$ , with  $\mathbf{r} = [r_1 \ r_2 \ \dots \ r_l]$  the vector of variables that condition the cFIR model.  $l$  should be kept to a low value, say below 3, owing to the curse of dimensionality (Hastie and Tibshirani 1990, pp. 83-84). When necessary, specific points related to the estimation of the two cFIR models introduced above will be discussed.

#### 3.1 Local polynomial estimates

The coefficient functions  $\beta_j(\mathbf{r})$  are estimated in a nonparametric framework, i.e. without assuming a shape for these functions. This is done by using local polynomial regression (Cleveland and Devlin 1988), for which the only assumption on the  $\beta_j$  coefficient functions is that they are sufficiently smooth for being locally approximated with polynomials. The estimation problem is reduced to locally fitting linear models at a number  $m$  of fitting points  $\mathbf{r}_{(i)}$ , where a given fitting point is defined by a pair of flux and time values (i.e. in



our case  $l = 2$ )

$$\mathbf{r}_{(i)} = [r_{(i),1} \ r_{(i),2} \ \dots \ r_{(i),l}]^\top, \quad i = 1, \dots, m \quad (19)$$

so that these  $m$  fitting points span the range of potential values on the various dimensions of  $\mathbf{r}$ . Defining these fitting points is preferably done by using some information on the distribution of  $\mathbf{r}$ . For the case of the cFIR models introduced above, this mainly concern the distribution of flux values, as hour values will be uniformly distributed.

Let us focus on the fitting point  $\mathbf{r}_{(i)}$  only. The local polynomial approximation  $\mathbf{z}_t$  of the vector of explanatory variables  $\mathbf{u}_t$  at  $\mathbf{r}_t = [r_{t,1} \ r_{t,2} \ \dots \ r_{t,l}]^\top$  is given by

$$\mathbf{z}_t = [u_{t,1}\mathbf{p}_d^\top(\mathbf{r}_t) \ \dots \ u_{t,k}\mathbf{p}_d^\top(\mathbf{r}_t) \ \dots \ u_{t,l}\mathbf{p}_d^\top(\mathbf{r}_t)]^\top \quad (20)$$

where  $\mathbf{p}_d(\mathbf{r}_t)$  corresponds to the  $d$ -order polynomial evaluated at  $\mathbf{r}_t$ . For instance if  $d = 2$ ,  $\mathbf{p}_2(\mathbf{r}_t)$  can be obtained as

$$\mathbf{p}_2(\mathbf{r}_t) = [1 \ r_{t,1} \ r_{t,2} \ r_{t,1}^2 \ r_{t,1}r_{t,2} \ r_{t,2}^2]^\top \quad (21)$$

In parallel, write

$$\boldsymbol{\phi}_{(i)} = \boldsymbol{\phi}(\mathbf{r}_{(i)}) = [\boldsymbol{\phi}_{(i),1}^\top \ \dots \ \boldsymbol{\phi}_{(i),k}^\top \ \dots \ \boldsymbol{\phi}_{(i),l}^\top]^\top \quad (22)$$

the vector of local coefficients at  $\mathbf{r}_{(i)}$ , where the element vector  $\boldsymbol{\phi}_{(i),k}$  is the vector of local coefficients related to the local polynomial approximation of the  $k$ -th explanatory variable, that is,  $u_{t,k}\mathbf{p}_d(\mathbf{r}_{(i)})$ .

The nonlinear cFIR is thus locally approximated at  $\mathbf{r}_{(i)}$  by the linear model

$$y_t = \mathbf{z}_t^\top \boldsymbol{\phi}_{(i)}, \quad \forall t \quad (23)$$

so that the problem of fitting the nonlinear cFIR model is converted in a number  $m$  of local linear models to be fitted, that is, one for each fitting point  $\mathbf{r}_{(i)}$ .

### 3.2 Offline estimation of the local coefficients

In an offline setting, a set of  $n$  observations for each of the time series is available and one then wants to estimate the local coefficients for this set of data. In such setting, with the one-step ahead prediction defined as the conditional expectation (cf. Section 2.2.3) the nonlinear cFIR model can be fitted by minimizing the sum of squared residuals over the set of observations, that is,

$$S(\boldsymbol{\beta}) = \sum_{t=1}^n \rho(y_t - \boldsymbol{\beta}^\top(\mathbf{r}_t)\mathbf{u}_t) \quad (24)$$

where  $\rho$  is a quadratic criterion, i.e. such that  $\rho(\epsilon) = \epsilon^2/2$ .

Then, if focusing on a given fitting  $\mathbf{r}_{(i)}$ , one can estimate the vector of corresponding local coefficients, that is,  $\boldsymbol{\phi}_{(i)}$ , by using weighted least-squares. The estimate  $\hat{\boldsymbol{\phi}}_{(i)}$  is then given

by

$$\hat{\phi}_{(i)} = \arg \min_{\phi_{(i)}} S(\phi_{(i)}) = \arg \min_{\phi_{(i)}} \sum_{t=1}^n w_{t,(i)} \rho(y_t - \mathbf{z}_t^\top \phi_{(i)}) \quad (25)$$

where the weights  $w_{t,(i)}$  are assigned by a Kernel function of the following form

$$w_{t,(i)} = T \left( \prod_{k=1}^l \frac{|r_{t,k} - r_{(i),k}|_k}{\hat{h}_{(i),k}(\alpha_k)} \right) \quad (26)$$

In the above,  $|\cdot|_k$  denotes the chosen distance on the  $k^{\text{th}}$  dimension of  $\mathbf{r}$ . For the cFIR models introduced in Section 2.2, one would choose an Euclidian distance on the dimension of the flux values and a polar distance on the dimension of hour values.

In parallel in (26),  $\hat{h}_{(i),k}$  is the bandwidth for that particular fitting point  $\mathbf{r}_{(i)}$  and for the  $k^{\text{th}}$  dimension of  $\mathbf{r}$ . Whatever the dimension considered,  $\hat{h}_{(i),k}$  is determined by using a nearest-neighbor principle (Nielsen et al. 2000). For a chosen proportion  $\alpha_k$ , the bandwidth  $\hat{h}_{(i),k}(\alpha_k)$  is such that

$$\alpha_k = \int_{\mathbb{D}_{(i),k}} f_{r_k}(v) dv \quad (27)$$

where  $\mathbb{D}_{(i),k} = \{v \in \mathbb{R} \mid |v - r_{(i),k}|_k < \hat{h}_{(i),k}\}$  defines the neighborhood of  $\mathbf{r}_{(i)}$  on the  $k^{\text{th}}$  dimension of  $\mathbf{r}$ , while  $f_{r_k}$  denotes the density function of the  $r_k$  values. In practice,  $f_{r_k}$  is replaced by the empirical distribution function of the available data.

Finally in (26),  $T$  is defined as the tricube function, i.e.

$$T : v \in \mathbb{R}^+ \rightarrow T(v) \in [0, 1], \quad T(v) = \begin{cases} (1 - v^3)^3, & v \in [0, 1] \\ 0, & v > 1 \end{cases} \quad (28)$$

as introduced and discussed by e.g. Cleveland and Devlin (1988).

Denote by  $\mathbf{A}$  the data matrix such that its  $t^{\text{th}}$  row is  $\mathbf{z}_t^\top$ , i.e. the transpose of the local polynomial approximation of  $\mathbf{u}_t$ . In parallel, write  $\Sigma$  the diagonal matrix for which the  $t^{\text{th}}$  element on the diagonal corresponds to the weight  $w_{t,(i)}$  to assign to the data point  $\mathbf{r}_t$ . The solution of (25) is then straightforwardly given by

$$\hat{\phi}_{(i)} = (\mathbf{A}^\top \Sigma \mathbf{A})^{-1} \mathbf{A}^\top \Sigma \mathbf{y} \quad (29)$$

where  $\mathbf{y}$  is the vector of observations for the time-series  $\{y_t\}$ .

The elements of  $\beta_{(i)}$  are finally obtained with

$$\hat{\beta}_{(i)} = \hat{\beta}(\mathbf{r}_{(i)}) = \mathbf{p}_d^\top(\mathbf{r}_{(i)}) \hat{\phi}_{(i)}, \quad i = 1, \dots, m \quad (30)$$

And, for a given data point  $\mathbf{r}_t$ , the corresponding coefficient functions  $\hat{\beta}(\mathbf{r}_t)$  are obtained by linear-type interpolation. For instance, if  $l = 2$ , it is done by using bilinear interpolation of the coefficient function values at the four fitting points forming the smallest surface that covers  $\mathbf{r}_t$ .

### 3.3 Online estimation of the local coefficients

For real-world applications, one does not want to consider the whole set of available observations for estimating the local coefficients every time new observations become available. Instead, for this online setting, one aims at tracking the local coefficients by using a recursive formulation of the estimation method. In addition, such recursive formulation can allow for an exponential forgetting of old observations, which leads to the model being adaptive with respect to long-term variations in the process characteristics. From now on, it is considered that at time  $n$  a set of  $n$  past observations is available for each time-series, and thus that the dataset grows as time increases.

First, let us introduce the objective function to be minimized at each time step  $n$ , which consists a modified version of that given by (24)

$$S_n(\boldsymbol{\beta}) = \sum_{t=1}^n \lambda^{n-t} \rho(y_t - \boldsymbol{\beta}^\top(\mathbf{r}_t)\mathbf{u}_t) \quad (31)$$

where  $\lambda$  is the forgetting factor that permits an exponential forgetting of past observations. For a given  $\lambda$ ,  $\lambda \in [0, 1[$ , the effective number of observations  $n_\lambda$  is

$$n_\lambda = 1 + \lambda + \lambda^2 + \dots = \frac{1}{1 - \lambda} \quad (32)$$

Denote by  $\hat{\phi}_{n,(i)}$  the estimate of the local coefficients for the fitting point  $\mathbf{r}_{(i)}$  at time  $n$ . Then, the objective function to be minimized for estimating the local coefficients at  $\mathbf{r}_{(i)}$  and at time  $n$  writes

$$S_n(\phi_{n,(i)}) = \sum_{t=1}^n \Lambda_{n,(i)}(t) w_{t,(i)} \rho(y_t - \mathbf{z}_t^\top \phi_{n,(i)}) \quad (33)$$

where  $\phi_{n,(i)}$  is related to  $\boldsymbol{\beta}_n(\mathbf{r}_{(i)})$  following a relation equivalent to (30). In parallel,  $\Lambda_{n,(i)}$  is the function that permits exponential forgetting of past observations, i.e.

$$\Lambda_{n,(i)}(t) = \begin{cases} \lambda_{n,(i)}^{\text{eff}} \Lambda_{n-1,(i)}(t-1), & 1 \leq t \leq n-1 \\ 1, & i = n \end{cases} \quad (34)$$

In the above definition,  $\lambda_{n,(i)}^{\text{eff}}$  is the effective forgetting factor for the fitting point  $\mathbf{r}_{(i)}$  which permits to account for the weighting in the formulation of (33). It follows the definition given by [Nielsen et al. \(2000\)](#), which tells that  $\lambda_{n,(i)}^{\text{eff}}$  is a function of  $w_{n,(i)}$  such that

$$\lambda_{n,(i)}^{\text{eff}} = 1 - (1 - \lambda)w_{n,(i)} \quad (35)$$

where  $\lambda$  is the classical user-defined forgetting factor,  $\lambda \in [0, 1[$ . This effective forgetting factor ensures that old observations are downweighted only when new information is available. This will be further explained in a following part of the present Paragraph. By using this exponential forgetting scheme,  $n_\lambda$  as given by (32) consists a lower bound on the effective number of observations ([Nielsen et al. 2000](#)).

The local coefficients  $\hat{\phi}_{n,(i)}$  at time  $n$  for model (23) are then given by

$$\hat{\phi}_{n,(i)} = \arg \min_{\phi_{(i)}} S_n(\phi_{(i)}) = \arg \min_{\phi_{(i)}} \sum_{t=1}^n \Lambda_{n,(i)}(t) w_{t,(i)} \rho(y_t - \mathbf{z}_t^\top \phi_{(i)}) \quad (36)$$

The recursive formulation for an adaptive estimation of the local coefficients  $\hat{\phi}_{n,(i)}$  (and therefore of  $\hat{\beta}_{n,(i)}$ , by using Equation (30) at each time-step) leads to the following three-step updating procedure at time  $n$ :

$$\epsilon_{n,(i)} = y_n - \mathbf{u}_n^\top \hat{\beta}_{n-1,(i)} \quad (37)$$

$$\hat{\phi}_{n,(i)} = \hat{\phi}_{n-1,(i)} + \epsilon_{n,(i)} w_{n,(i)} (\mathbf{R}_{n,(i)})^{-1} \mathbf{z}_n^\top \quad (38)$$

$$\mathbf{R}_{n,(i)} = \lambda_{n,(i)}^{\text{eff}} \mathbf{R}_{n-1,(i)} + w_{n,(i)} \mathbf{z}_n \mathbf{z}_n^\top \quad (39)$$

where  $\lambda_{n,(i)}^{\text{eff}}$  is again the effective forgetting factor. One sees that when the weight  $w_{n,(i)}$  equals 0 (thus meaning that the local estimates should not be affected by the new information), then we have  $\hat{\phi}_{n,(i)} = \hat{\phi}_{n-1,(i)}$  and  $\mathbf{R}_{n,(i)} = \mathbf{R}_{n-1,(i)}$ . This confirms the role of the effective forgetting factor, that is to downweight old observations, but only when new information is available.

For initializing the recursive process, the matrices  $\mathbf{R}_{0,(i)}$ ,  $i = 1, \dots, m$ , can be chosen as

$$\mathbf{R}_{0,(i)} = \delta \mathbf{I}, \quad \forall i \quad (40)$$

where  $\delta$  is a small positive number and  $\mathbf{I}$  is an identity matrix of appropriate size. Note that this size, which corresponds to the number of coefficients to be estimated, is equal to the order of the chosen model in Equation (23) times the order of the polynomials used for local approximation. In parallel, the coefficient functions are initialized with a vector of zeros, or alternatively from a best guess on the target regression.

### 3.4 Regularization for a better generalization ability

The cFIR models (7) and (9) are originally designed for performing one-step ahead prediction. However, it is used here for multi-step ahead forecasting purposes with flux predictions as input. One should then try not to amplify the error in flux forecasts when passed through the model. This can be done by insuring that we work with a ‘smooth’ model. In addition, using recent data for fitting the model does not ensure an optimal performance when consequently used for predicting on new and independent data. This ability of performing well with new and independent data is referred to as the generalization ability of the model (see e.g. Stone (1974)). For these two reasons, we propose here a regularized version of the estimation method described in the above Paragraph.

Several approaches may be considered for regularization in recursive least squares methods. They are widely used for ill-conditioned numerical problems, for avoiding overfitting when training neural networks (Leung et al. 1999, Sjöberg & Ljung 1995) or more generally for estimation in nonlinear systems (Bishop 1995, Johansen 1997). Here, the type of regularization applied is that known as Tikhonov regularization (Tikhonov and Arsenin 1977). It consists in adding a penalty term related to the norm of the coefficients (or of

their derivatives) in the loss function to be minimized for model fitting.

For the case of the offline estimation of the model coefficients, it is well known that adding Tikhonov regularization makes that equation (29) becomes

$$\hat{\phi}_{(i)} = \left( \mathbf{A}^\top \Sigma \mathbf{A} + \mu \mathbf{I} \right)^{-1} \mathbf{A}^\top \Sigma \mathbf{y} \quad (41)$$

where  $\mathbf{I}$  is an identity matrix of appropriate size, and where  $\mu$  is the regularization parameter that permits to control the trade-off between the minimization of the fitting errors and the norm of the model estimates. One thus sees that Tikhonov regularization consists in penalizing the diagonal elements of the inverse covariance matrix.

If going to the case of online estimation of the cFIR model coefficients, the loss function to be minimized at each time  $t$  can be reformulated as

$$\tilde{S}_n(\phi_{n,(i)}) = \frac{\mu}{1-\lambda} \phi_{n,(i)}^\top \phi_{n,(i)} + \sum_{i=1}^n \Lambda_{n,(i)}(t) w_{n,(i)} \rho(y_t - \mathbf{z}_t^\top \phi_{n,(i)}) \quad (42)$$

where  $\phi_{n,(i)}^\top \phi_{n,(i)}$  represents the quadratic norm of the model estimates. We restrict ourselves to the case for which  $\lambda < 1$ . The regularization parameter  $\mu$  is multiplied here by  $(1-\lambda)^{-1}$ , which corresponds to the effective number of observations for this loss function formulation. By doing so,  $\mu$  represents the regularization load to be added to each observation accounted for in the loss function. The regularization is thus independent with the size of the virtual sliding window considered, which is in turn controlled by  $\lambda$ . If no exponential forgetting was used, a single parameter  $\mu$  would multiply the norm of the model estimates. Choosing  $\mu$  in such case would be an issue, as the result of the sum in (42) would increase as  $n$  increases, while  $\mu$  is not a function of  $n$ . Note that when regularization in RLS methods is considered in the literature, see e.g. (Hubing and Alexander 1991, Ismail and Principe 1997), it is always for the case  $\lambda = 1$  and with effect of the regularization fading as the size of the dataset increases. The aim of such regularization is mainly to control first adaptation steps after model initialization.

For the adaptive formulation of the loss function in (42), the recursive procedure for updating the cFIR model coefficients at each time step can be rewritten in order to account for regularization. In fact, the main difference with the classical updating procedure described above relates to the updating formula for the inverse covariance matrix  $\mathbf{R}_{n,(i)}$ . One indeed then works with a regularized inverse covariance matrix  $\tilde{\mathbf{R}}_{n,(i)}^*$ , which replaces  $\mathbf{R}_{n,(i)}$  in (38), and which is updated with

$$\tilde{\mathbf{R}}_{n,(i)} = \lambda_{n,(i)}^{\text{eff}} \tilde{\mathbf{R}}_{n-1,(i)} + w_{n,(i)} \mathbf{z}_n \mathbf{z}_n^\top + \frac{1 - \lambda_{n,(i)}^{\text{eff}}}{1 - \lambda} \mu \mathbf{I} \quad (43)$$

where  $\mathbf{I}$  is an identity matrix of appropriate size. The inverse covariance matrix  $\tilde{\mathbf{R}}_{(i),0}$  can be initialized with  $\delta = \mu$  in equation (40). Such a recursive scheme for the updating of the covariance matrices makes that its diagonal elements are always penalized with the same regularization parameter  $\mu(1-\lambda)^{-1}$  (as it is the case in (41)). The model estimates are still updated by applying (38) at each time step.

## 4 Case-studies and results

### 4.1 Description of the case-studies

The models and forecasting methodology described above are applied to the test case of the Roskilde district heating system. Roskilde is a city of around 250.000 inhabitants located 30 kilometers west of Copenhagen in Denmark. The district heating system has a unique heat production facility, and focus is given to three critical points on the distribution network. They correspond to a retirement home, the local hospital and the Viking museum. They are hereafter referred to as critical points 1, 2 and 3, respectively.

The available data consist in measurements of the water flux and temperature at the supply point, as well as measurements of the water temperature at the three critical points. They have a time resolution of 5 minutes. Hourly data are obtained by averaging the 5-minute measurements so that, for instance, the hourly temperature value at 01:00 is the average of all values between 00:05 and 01:00. If more than two measurements are missing for a given hour, the corresponding hourly value is considered as erroneous. The period for which measurements are available goes from August 16<sup>th</sup>, 2000, to December 12<sup>th</sup>, 2006. This translates to 55423 hourly values for each variables. For the specific case of critical point 2, the last 11000 data points are not considered owing to a suspicious behavior of the time-series, which is in turn due to the application of local control strategies of the water temperature at Roskilde hospital from 2005. The overall percentage of valid data for the three critical points are of 88.49, 86.28% and 94.56%, respectively.

The aim of the present exercise is to demonstrate the significantly higher performance of the proposed forecasting methodology in comparison with the state-of-the-art approach described in Section 2.1. But also, the structure of the proposed models, and the two rival approaches to the integration of the social behavior of the consumers, will allow us to discuss the time-delays in the network and the assumption such that the social behavior impacts (or not) the transfer function of the distribution network.

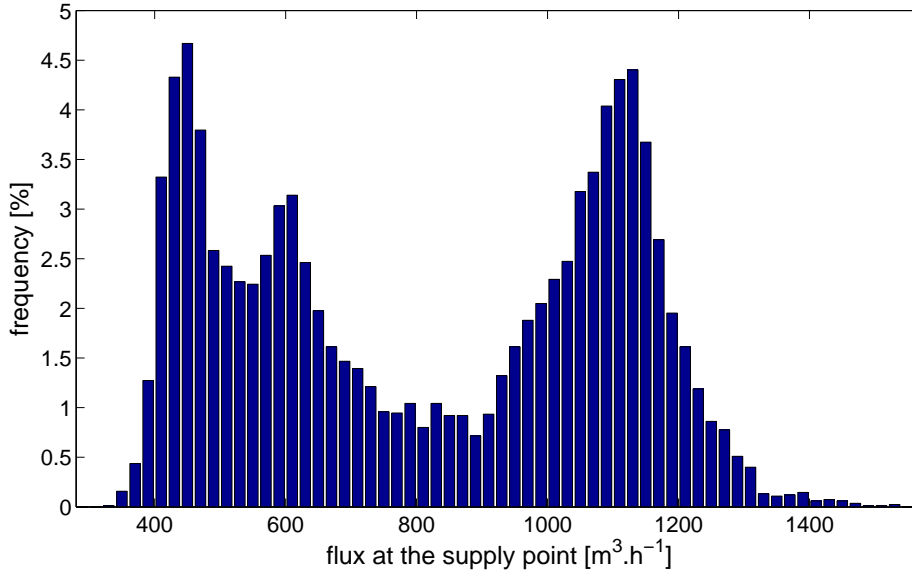
### 4.2 Setup and optimal fitting of the cFIR models

The cFIR models (7) and (9) comprise the central part of the forecasting methodology. For both models the set of indices that defines the past values of the water temperature at the supply point is such that  $S_j = \{1, 2, \dots, 10\}$ , i.e. the cFIR models rely on the last 10 hourly values of supply temperature.

Regarding local polynomial regression, one has to start by defining the order of the polynomials used for locally approximating the regression. Linear polynomial regression is used in order to limit the number of local coefficients to be estimated, and thus the related computational costs. In addition, it is necessary to set the fitting points and the proportions that defines the nearest-neighbor bandwidths. For both rigid and supple cFIR models, because of the yearly cycle in the flux values, it is chosen to use one year of flux values for obtaining a representative empirical distribution  $\hat{f}_x$  of the flux values. This distribution of flux values can be seen from Figure 1.

The fitting points along the flux dimension  $x_{(i)}$ ,  $i = 1, \dots, m_x$ , are then defined such that there is the same proportion of flux values in each of the intervals defined by two successive





**FIGURE 1:** One year distribution of flux values at the supply point. The bin size is set to  $25 \text{ m}^3 \cdot \text{h}^{-1}$ .

fitting points. This writes

$$\int_0^{x(i)} f_{\bar{x}}(v) dv = \frac{i-1}{m_x-1}, \quad i = 1, \dots, m_x \quad (44)$$

The number of fitting points mainly has an impact on the computational costs for model estimation, provided that  $m_x$  is set to a sufficient value so that local polynomial approximation is suitable. It has been found that the model fitting was not significantly better when having  $m$  above 11, and it has been chosen to use this value. The related bandwidth values  $\hat{h}_{(i),x}$ ,  $i = 1, \dots, m_x$  are obtained by applying the nearest-neighbor principle introduced in (27), parameterized by  $\alpha_x$ . Similarly, we have witnessed that the improvement in the model fitting was negligible when having  $\alpha_x$  higher than 0.4, and it is thus the value chosen in the following. For the specific case of the supple cFIR model, it is chosen to have four fitting points uniformly distributed over the range of daily hours. The local coefficients for the supple cFIR models will hence be estimated for 00:00, 6:00, 12:00 and 18:00. In order to have very smooth variations along this dimension of the supple cFIR, the related bandwidth is set to a large value ( $\alpha_h = 0.6$ ).

In parallel, concerning the recursive estimation method itself, one has to set the value of the forgetting factor  $\lambda$ , which defines the rate of forgetting of old observations, and controls the ability of the method to account for long-term variations in the process characteristics. Though, if choosing a value for  $\lambda$  that is too low, the fitting of the model will be very poor. Finally, a last parameter to consider is the regularization parameter  $\mu$ , defining the trade-off between model fitting and the smoothness of the model estimates. Our methodology for defining optimal values for these two parameters is to use one-fold cross validation. The first year of the available dataset is used as a training period, while the second year is seen as the validation set. Since it is considered that the cFIR model will be used in practice for generating temperature predictions up to 12-hour ahead, the criterion we choose to minimize on the validation set is the Root Mean Square Error (RMSE), averaged over this

**TABLE 1:** Results from the cross-validation procedure for the choice of the optimal forgetting factor  $\lambda$  and regularization parameter  $\mu$ . These results are for the rigid cFIR model (a) and supple cFIR model (b).

(a) rigid cFIR model			
critical point number	$\lambda$	$\mu$	mean RMSE [ $^{\circ}$ C]
1	0.98	0.00075	0.4851
2	0.98	0.0008	0.4902
3	0.9	0.1	0.8210

(b) supple cFIR model			
critical point number	$\lambda$	$\mu$	mean RMSE [ $^{\circ}$ C]
1	0.975	0.001	0.5014
2	0.98	0.0006	0.5304
3	0.9	0.1	0.8108

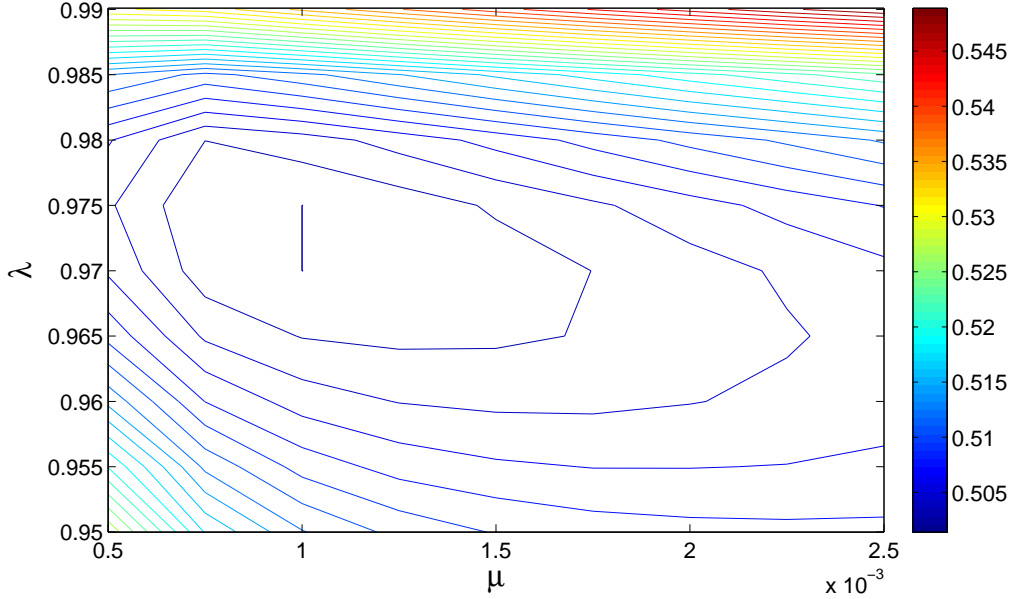
range of forecast horizons. For that purpose of model fitting, only flux measures are used.

The results from the cross-validation procedure are gathered in Table 1. For the three critical points, the optimal parameters  $\lambda$  and  $\mu$  are very similar if considering the rigid or supple cFIR models. For two out of the three critical points, the mean RMSE related to the optimal parameters is slightly higher for the supple cFIR model than for the rigid cFIR model. This is not the case for the third critical point. However, one sees that the value of the optimal forgetting factor is low (0.9), while the regularization parameter value is very high (0.1) in comparison with the other critical points. The mean RMSE also reaches a surprisingly large level in comparison with the two other critical points. This might indicate some abnormal behavior in the data, owing to e.g. the quality of the temperature measurements. One should note though that this critical point corresponds to the Viking museum, which has the lowest contribution to the heat consumption. This might introduce some different behavior in the data that cannot be captured by the proposed cFIR models.

A nice feature of the cFIR models is the low sensitivity to the choice of the two parameters  $\lambda$  and  $\mu$ . For instance, Figure 2 shows the variation of the average RMSE on the validation set, for the supple cFIR model applied to critical point 1, as a function of  $\lambda$  ( $y$ -axis) and  $\mu$  ( $x$ -axis). These variations are described by a contour plot, with 30 level lines related to uniformly distributed mean RMSE values. The minimum average RMSE is obtained for  $(\lambda, \mu) = (0.975, 0.001)$ .

From Figure 2, one sees that the 3-dimensional surface describing the variations is a smooth and convex surface. The convex nature of this surface makes that there is a unique  $(\lambda, \mu)$  combination that minimizes the mean RMSE value on the cross-validation set. Also, the fact that this surface is smooth demonstrates the low sensitivity of the performance of the cFIR models to the choice of the parameters  $\lambda$  and  $\mu$ .

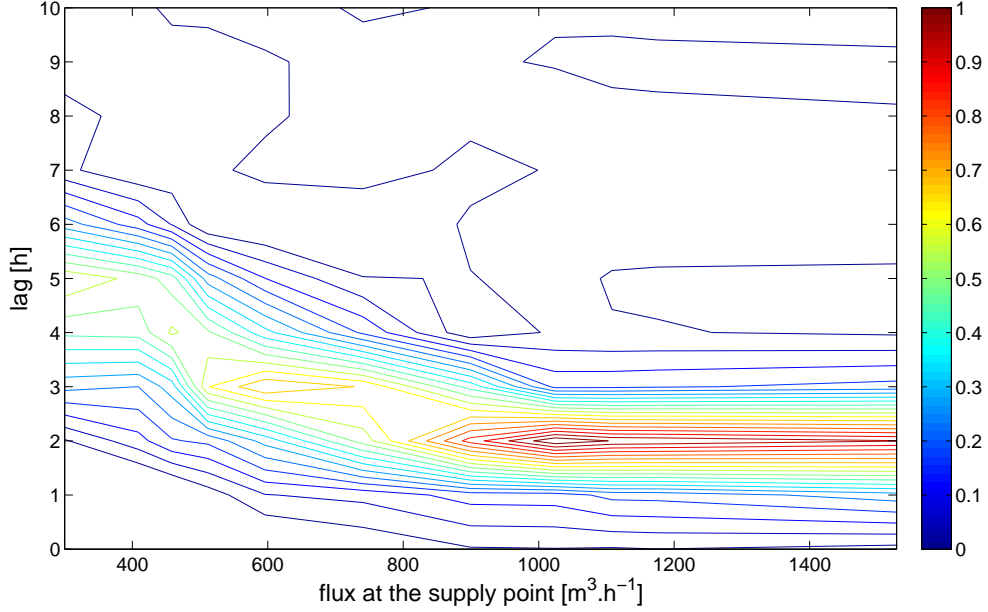
In a last stage, we depict the estimated cFIR model related to the transfer function of the district heating system between the supply point and critical point 1. The contour plot of Figure 3 indeed gives the amplitude of the coefficient functions in the cFIR model, as a function of both the water flux and the lag of the water temperature at the supply point. For



**FIGURE 2:** Mean Root Mean Square Error [ $^{\circ}\text{C}$ ] over the validation set as a function of both the forgetting factor  $\lambda$  (y-axis) and the regularization parameter  $\mu$  (x-axis). These results are for the supple cFIR model and for critical point 1.

high flux values, the transfer function of the district heating network between the supply point and critical point 1 is highly concentrated: the water temperature at critical point 1 is almost only determined by that at the supply point 2 hours before. The coefficient function takes a value of almost 1 for this lag, while it sharply decreases for shorter and longer lags. One can observe such behavior for flux values down to  $1000\text{m}^3\cdot\text{h}^{-1}$ . For lower values, the transfer function is less concentrated and more lagged values of the water temperature at the supply point contribute to determining that at critical point 1. This may be explained by a different mixing of the water in the pipe owing to a lower flux. In addition, it has been noticed (for both cFIR models and for the three critical points) that the offset term in the cFIR tends to decrease as the flux values are lower, indicating higher temperature losses.

The interest of Figure 3 is also that it permits to better appraise the physical behavior of the district heating system between the supply point and a given critical point. The influence of the flux on the losses has already been mentioned above. In addition to this, one can also visualize the influence of the flux at the supply point on the time delays in the network. It varies here between 2 and 5 hours when going from the highest to lowest flux values. The time-delay variations appear to be nonlinear, though this may come from some numerical artifact. Indeed, the same regularization parameter  $\mu$  is used for all fitting points. However, since an optimal regularization parameter (for a linear model) is related to the variance and the norm of model estimates (see e.g. (Golub et al. 2000, Wang and Chow 1989)), the optimal  $\mu$  for the estimation of the coefficient functions in the cFIR should also be a function of the fitting point considered. This is because both the variance and the norm of the coefficient functions will vary depending on the fitting point. The optimal and local tuning of the regularization parameter  $\mu$  should be further investigated in the future. Note that for the case of the supple cFIR models, one could also visualize contour plots such as that of Figure 3 for different hours of the day. Since we have not witnessed highly significant variations in the ‘shape’ of the transfer functions through



**FIGURE 3:** Contour plot of the amplitude of the coefficient functions of the rigid cFIR model at the end of the validation set for critical point 1. The x-axis gives the flux at the supply point (in  $m^3 \cdot h^{-1}$ ) while the y-axis gives the lag of the transfer function in hours.

the day, this point is not further discussed here.

### 4.3 Out-of-sample evaluation of the cFIR models

The remaining of the dataset is seen as an out-of-sample evaluation set, for which the application of the forecasting methodology has to mimic operational conditions, so that observed performance would be representative. The evaluation set consists of 33900 data points ( $\sim$  four years) for critical points 1 and 3, and of 22500 data points (slightly less than three years) for critical point 2. The cFIR models estimated for the three critical points permit to describe the transfer function between the supply point and each of these critical points. For control purposes in operational conditions, these models can serve for simulating what would be the temperature at a given critical point for the following hours, depending on the chosen control strategy (on both flux and water temperature variables) at the supply point. Therefore, for the out-of-sample evaluation of the cFIR models, the measurements of both the flux and the water temperature at the supply point are used as input to the models.

Several criteria are considered for evaluating the prediction performance. All these criteria are calculated on a per-horizon basis, since it is expected that the prediction performance deteriorates as the lead time gets further. The bias, which equals the mean of all prediction errors, corresponds to the systematic part in the prediction error. Then, the Mean Absolute Error (MAE) gives the average deviation in absolute value between forecasts and measures. Finally, the Root Mean Square Error (RMSE) is the average of squared errors, thus giving more weight to large prediction errors. This last error measure is used as the main criterion when evaluating the prediction performance, since it relates to the

quadratic loss function used for model estimation (and parameter selection through the cross-validation procedure). Note that the range of temperature variations over the whole dataset are of 31.26°C, 28.58°C and 30.53°C for critical points 1, 2 and 3, respectively. A final criterion that will be used is the Mean Absolute Percentage Error (MAPE), for which absolute errors are divided at each time step by the measured temperature. The full set of results from the evaluation are gathered in Tables 5, 6, and 7 in the Appendix.

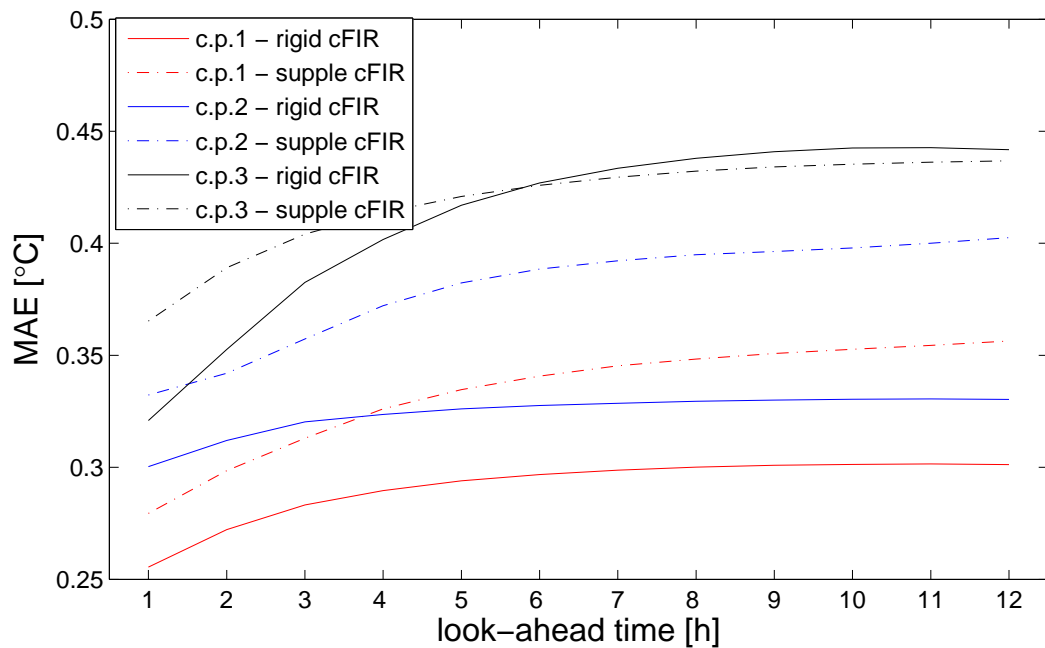
Figure 4 summarizes the evaluation of the cFIR models with the MAE and RMSE error measures calculated as a function of the look-ahead time. There is indeed a slight increase of the prediction error as the lead time gets further for the first 2 critical points, while this increase is more pronounced for the third one. In addition, the average level of prediction error is significantly higher for this last critical point, as it was also the case when fitting the cFIR models from the cross-validation procedure. The MAE averaged over the forecast length ranges between 0.291 and 0.597°C depending on the cFIR models and the critical point considered. For the first 2 critical points, there is significant difference in the accuracy of predictions generated with the supple and rigid cFIR models, with a clear advantage for the latter ones. However for the third critical point their forecast accuracy are much more similar, with this time an advantage for the supple cFIR models for horizon further than 5-6 hours ahead.

#### 4.4 Sensitivity to the flux values used as input to the cFIR models

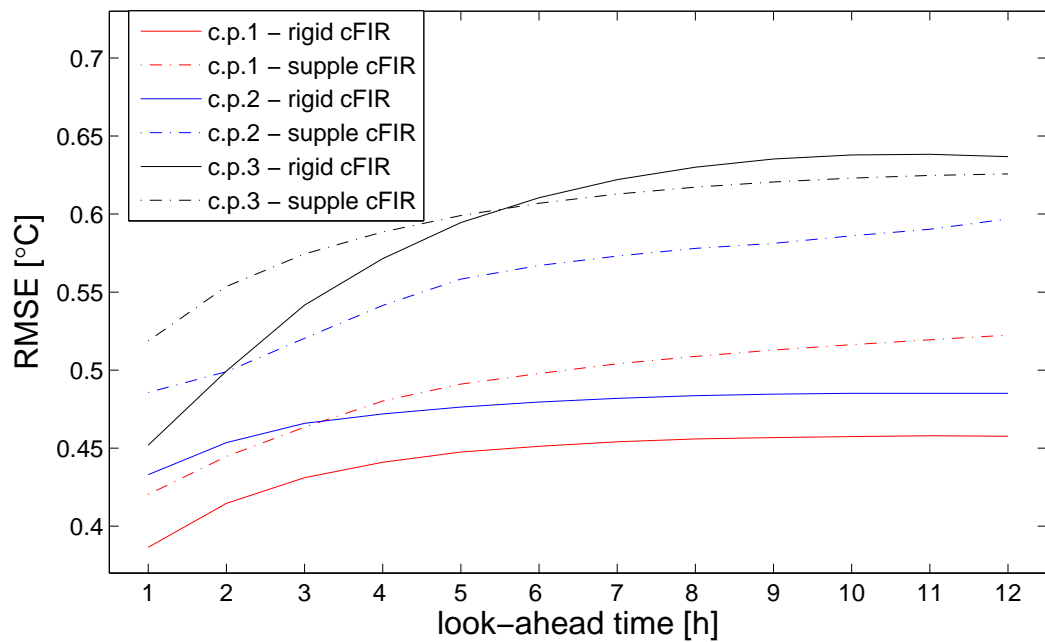
In this Section, a sensitivity analysis on the performance of the proposed forecasting methodology is carried out. Especially, focus is given to the sensitivity of this performance depending on the flux values used as input to the cFIR models. It is then imagined here that in operational conditions a simple model is used for predicting the flux at the supply point for the following hours. The accuracy of forecasts obtained with the autoregressive model (18) is assessed in a first part. Subsequently, instead of using flux measurements or flux scenarios as input to cFIR models multi-step ahead prediction, they are fed with the forecasts obtained by using the autoregressive model. The influence on the performance of the resulting temperature predictions is discussed.

##### 4.4.1 Flux prediction at the supply point

In a first stage, focus is given to the issue of multi-step ahead prediction of the flux at the supply point. For that purpose, model (18) is fitted by using a RLS estimation method with exponential forgetting. The forecast length is set to 12 hours, and the order of the AR model is chosen to be 24 hours. One-fold cross validation is performed for selecting an optimal forgetting factor, the optimality criterion being defined as the Root Mean Square Error (RMSE) averaged over the forecast length. The first 1000 points are used as a learning period, and the following 3000 (approximately 4 months) serve for the cross-validation procedure mentioned above. To give the reader an idea on the variations of the flux at the supply point, the minimum and maximum values over the dataset are of 300m<sup>3</sup>.h<sup>-1</sup> and 1528m<sup>3</sup>.h<sup>-1</sup>, respectively. The trend is to witness a yearly cycle with lower flux values in summer (owing to a lower demand) and higher flux values in winter. It would have then seemed more appropriate to use a whole year for cross-validation instead of the considered 4 months. However, as mentioned in Section 2.2, our aim is not here to have the best possible performance on flux prediction, but more to have realistic predictions, the performance



(a) MAE as a function of the look-ahead time.



(b) RMSE as a function of the look-ahead time.

**FIGURE 4:** Out-of-sample evaluation of the forecast performance of both the rigid and suppl cFIR models for the three critical points. Flux measurements are used as input to the cFIR models. The forecast performance is evaluated with MAE and RMSE error measures as a function of the look-ahead time up to 12-hour ahead.



of which will give a lower bound on the quality of flux predictions that would be used in future applications. Finally, the remaining data points (approximately 55400) are used for the out-of-sample evaluation of the proposed flux prediction method. Out of this evaluation set, 84.22% of valid data can be used for forecast evaluation. Error measures such as bias, MAE, RMSE (in  $\text{m}^3.\text{h}^{-1}$ ), and MAPE (in %) are considered. The results are gathered in Table 2.

**TABLE 2:** *Some prediction accuracy measures related to flux forecasting in the Roskilde district heating system. These measures are the bias, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE). The first three error measures are expressed in  $\text{m}^3.\text{h}^{-1}$  while the latter one is in percentage of predicted output.*

horizon [h]	bias [ $\text{m}^3.\text{h}^{-1}$ ]	MAE [ $\text{m}^3.\text{h}^{-1}$ ]	RMSE [ $\text{m}^3.\text{h}^{-1}$ ]	MAPE [%]
1	0.011	17.790	25.233	2.31
2	0.006	29.996	42.036	3.82
3	0.036	37.683	52.326	4.80
4	0.031	42.196	58.302	5.38
5	-0.015	44.913	62.061	5.73
6	-0.052	46.554	64.442	5.94
7	-0.108	47.586	65.991	6.07
8	-0.158	48.219	66.955	6.14
9	-0.187	48.644	67.628	6.20
10	-0.210	49.057	68.263	6.25
11	-0.232	49.437	68.845	6.30
12	-0.267	49.864	69.514	6.35

Comparing the magnitude of the bias with that of the RMSE, one appraises that predictions are relatively unbiased whatever the look-ahead time. Owing to the auto-regressive structure of the model chosen for prediction, predictions for a look-ahead time  $k$  are necessarily produced by feeding the model with predictions for horizons up to  $k - 1$ . This makes that even if model fitting is based on minimizing the average quadratic error over the forecast length for a better generalization ability, prediction errors still sum up as the look-ahead time increases. This increase is here highly significant for the first forecast horizons, while the level of prediction error stabilizes for further look-ahead times. The maximum MAPE, which corresponds to 12-hour ahead forecasts, reaches 6.35%.

#### 4.4.2 Performance of cFIR models fed with flux predictions

Instead of using the cFIR models for simulations purposes, as it has been done in Section 4.3, they are used here for genuine prediction purposes. It is not assumed that one could feed cFIR models with flux measurements for multi-step ahead predictions. The flux forecasts obtained by the autoregressive model (18), the performance of which have been discussed in the above Paragraph, are used instead.

The evaluation of the performance of the multi-step ahead predictions resulting from this approach are gathered in Tables 5, 6, and 7 in the Appendix. In parallel, Figure 5 summarizes the evaluation of the cFIR models with the MAE and RMSE error measures calculated as a function of the look-ahead time. It can then be directly compared to the results

of Figure 5 in order to quantify the loss in forecast accuracy owing to the use of flux values that are not the true values for feeding the cFIR models.

The range of variations of the MAE and RMSE error measures as a function of the look ahead time appears to be similar to that observed in Figure 5, i.e. in the case for which flux measurements input cFIR models. The increase in error measures as the lead time gets further is slightly sharper for the case of critical points 1 and 2, but not for the third one. The MAE averaged over the forecast length goes here from 0.318 and 0.590°C depending on the cFIR models and the critical point considered. Because of the smooth nonlinear variations of the transfer function of the network (if seen as a function of the flux at the supply point), the errors in flux predictions are actually dampened when passed through the cFIR models. Obviously, using more advanced methods for flux prediction (such as that described in Section 2.2.3) should permit to have more accurate temperature forecasts, but the impact on the gain in accuracy would be limited.

Note that while the loss in prediction accuracy is much more significant for the rigid than for the supple cFIR models. The supple cFIR models still appear to be globally superior, though the difference in prediction accuracy between the two types of cFIR models is now very low.

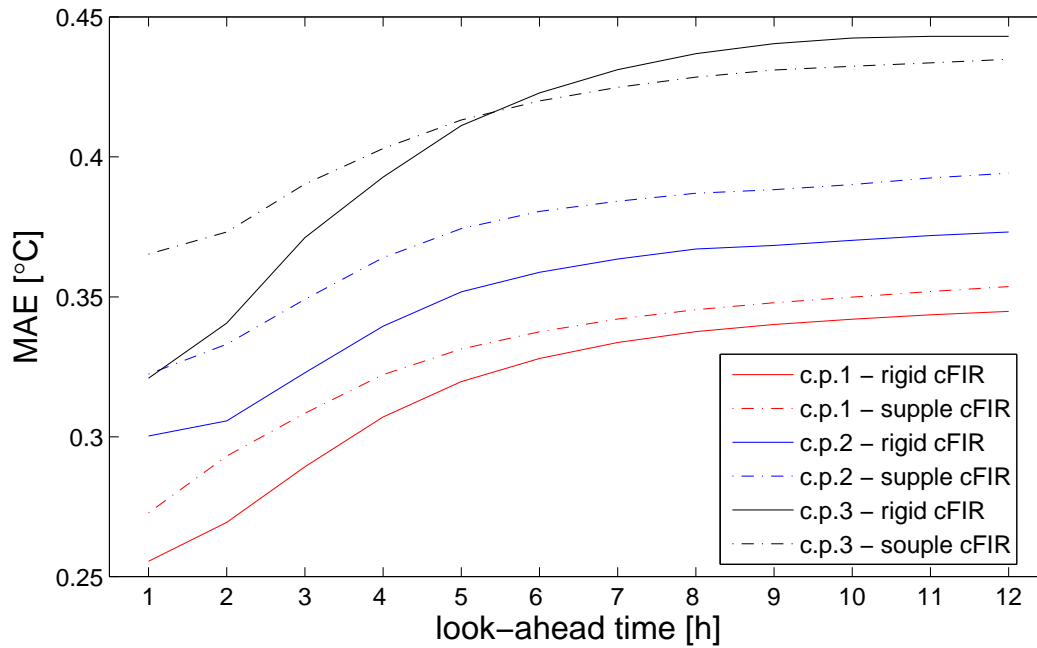
#### 4.5 Comparison with the state-of-the-art approach

In a last part of this study, focus is given to comparing the introduced forecasting methodology to the state-of-the-art statistical approach. This one is based on the lagged transfer function of model (1). As explained in Section 2.1, the lag  $\tau$  in this model is chosen as that which maximizes the correlation between the time series  $\{y_t\}$  and  $\{u_{t-\tau}\}$ . A single time-delay is considered over the whole dataset. The chosen lag for each of the critical points, as well as the corresponding correlation values are gathered in Table 3. Significant differences in correlation values for the 3 critical points can be noticed. The fact that the correlation is higher and the lag shorter for critical point 2 indicates that this critical point has a larger share of the heat consumption. Note that observing a higher correlation does not mean that one should expect a higher forecast accuracy of the lagged transfer function approach. This will indeed be shown in the following.

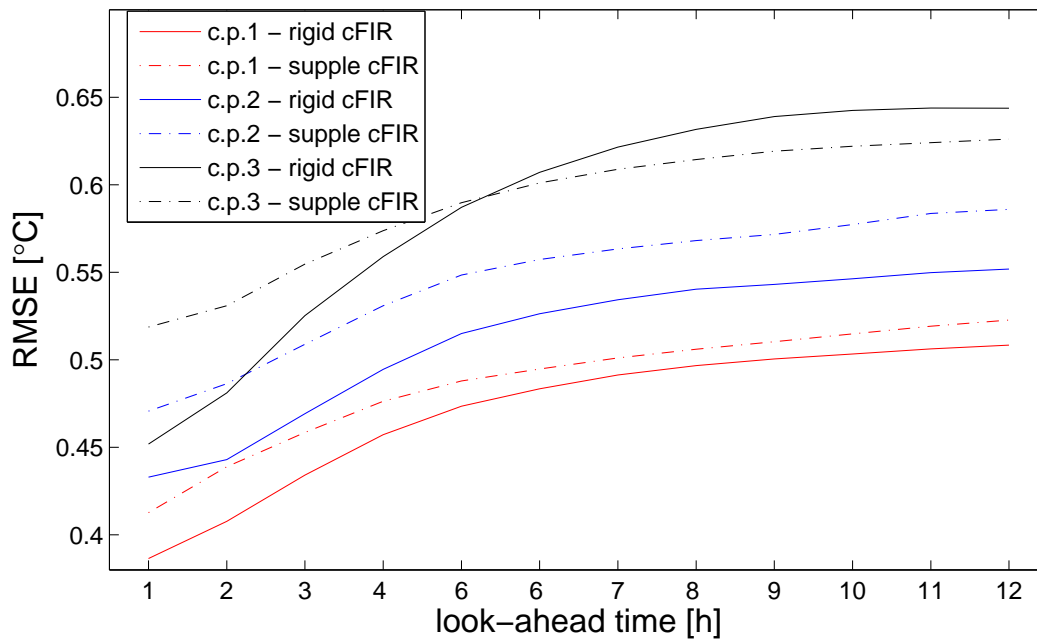
**TABLE 3:** Characteristics of the lagged autoregressive models used for temperature predictions. The lag of each model is that which maximizes the correlation between the lagged values of the water temperature at the supply point and the temperature measurement at each critical point. The forgetting factor is obtained from a one-fold cross-validation procedure.

critical point	lag [h]	correlation	$\lambda$
1	2	0.421	0.988
2	1	0.814	0.994
3	2	0.742	0.98

The model coefficients are adaptively estimated with a RLS method with exponential forgetting. The optimal value of the forgetting factor is obtained from a one-fold cross-validation procedure, similar to that employed for optimal tuning of the cFIR models or of the autoregressive model used for flux predictions. In order to have a fair comparison between cFIR models and the lagged transfer function models, we also use here the sec-



(a) MAE as a function of the look-ahead time.



(b) RMSE as a function of the look-ahead time.

**FIGURE 5:** Out-of-sample evaluation of the forecast performance of both the rigid and supplé cFIR models for the three critical points. Flux predictions are used as input to the cFIR models. The forecast performance is evaluated with MAE and RMSE error measures as a function of the look-ahead time up to 12-hour ahead.

ond year of the data as a validation set, and the remaining of the dataset for out-of-sample evaluation of the forecast performance. The autoregressive and cFIR-based approaches are evaluated on the same period and same data. The optimal forgetting factors are also given in Table 3. The level of forgetting is again more pronounced for critical point 3.

The full evaluation of the accuracy of the multi-step ahead predictions for the 3 critical points, resulting from the use of the lagged transfer function model of the state of the art, is summarized in Tables 11, 12, and 13 in the Appendix. In addition, Figure 6 depicts the evolution of the MAE and RMSE error measures as a function of the look-ahead time for this set of test cases.

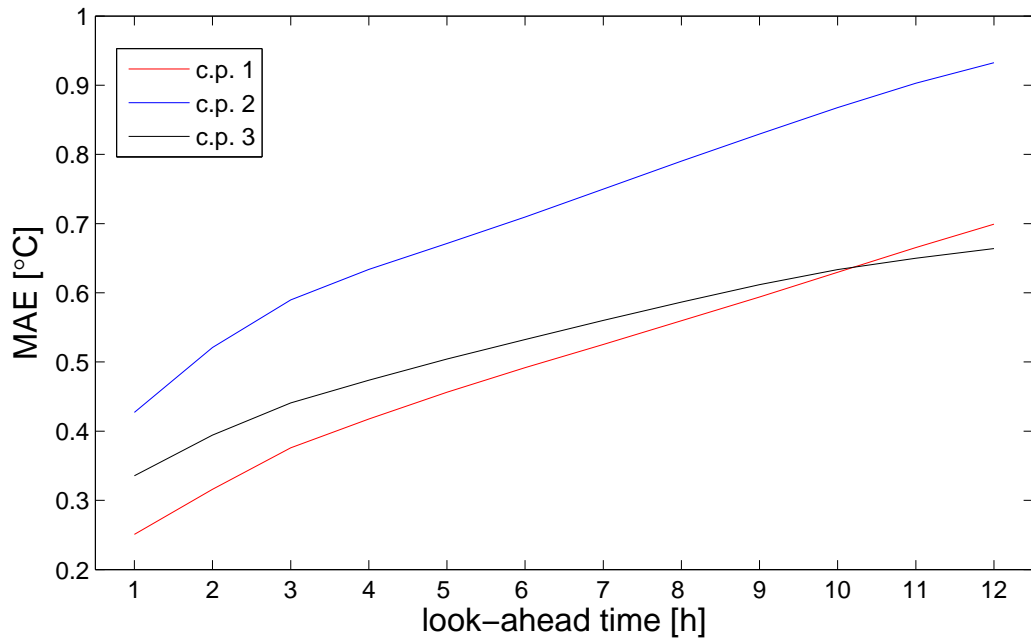
Tables 11, 12, and 13 show that the bias increases with the look-ahead time. It reaches (for 12-hour ahead prediction) levels that are much higher than that of forecasts generated with the cFIR models. Also, the MAE and RMSE error measures increase almost linearly with the forecast horizon (cf. Figure 6), thereby illustrating the comment in Section 2.1 such that forecast errors would sum up as the lead time gets further, owing to the autoregressive component of model (1), thus dramatically affecting forecast accuracy.

The prediction performance is higher for critical points 1 and 3, while for the case of cFIR models it was actually higher for critical points 1 and 2. There is not obviously critical points where it is easier or more difficult to predict. It may depend on the forecasting approach considered. The MAE for critical points 1 and 3 ranges between 0.3 and 0.7°C while it goes from 0.4 to 0.9°C for the case of critical point 2. If one compares with the evaluation results shown in Figures 4 or 5 for cFIR models, one sees that there is a dramatic difference between the levels of prediction accuracy for the two approaches. For appraising the reduction in prediction error when going from the state-of-the-art approach to that based on cFIR models, the improvement with respect to the RMSE criterion is calculated for each critical point and for each cFIR type. This improvement is given by the decrease in RMSE divided by the RMSE for the lagged transfer function models. The improvement values are gathered in Table 4 as a function of the look-ahead time.

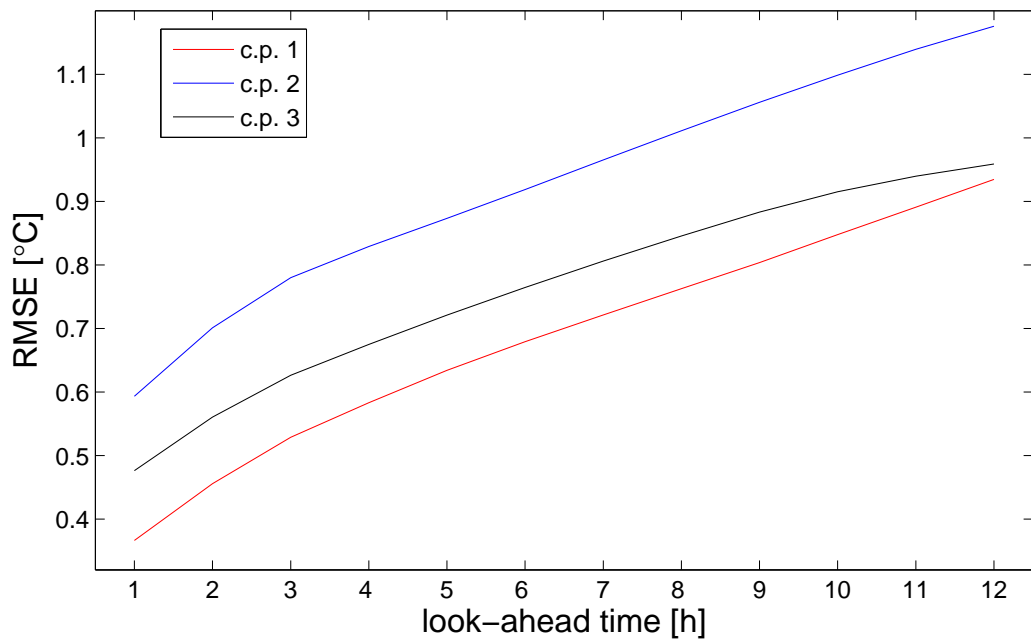
These improvement values increase as the lead time gets further. For 1-hour ahead forecasts, some of them are negative indicating that for such short horizon the use of the state-of-the-art approach may be beneficial. However, improvement values reach a very high level rapidly, with a decrease in RMSE up to 53% for 12-hour ahead predictions for critical point 2. The improvement is slightly less for the two other critical points, but the forecasting method based on cFIR models clearly outperforms the state-of-the-art approach based on the lagged transfer function model.

## 5 Concluding remarks

Models for the prediction of the temperature at critical points of district heating systems are paramount for the heat suppliers to make optimal decisions on the water temperature at the supply point. This is because the decision-making methodologies are based on model predictive control. It is thus expected that improvements in the accuracy of temperature forecasts at critical points will significantly improve the control decisions on supply temperature, and thus lower the operational production costs of heat suppliers. The forecasting methodology introduced in the present paper contributes to reaching a higher accuracy of such temperature forecasts. An evaluation of the performance of this methodology has



(a) MAE as a function of the look-ahead time.



(b) RMSE as a function of the look-ahead time.

**FIGURE 6:** Out-of-sample evaluation of the forecast performance of the lagged transfer function model of the state of the art. The forecast performance is evaluated with MAE and RMSE error measures as a function of the look-ahead time up to 12-hour ahead.

**TABLE 4:** *Improvements with respect to the RMSE criterion achieved by the genuine predictions from cFIR models in comparison with forecasts obtained from the state-of-the-art approach. Improvements are given for each critical point, both types of cFIR model, and are expressed in %.*

(a) rigid cFIR model

horizon [h]	critical point 1	critical point 2	critical point 3
1	-5.5146	26.9938	5.0830
2	10.5332	36.7955	14.1353
3	17.9085	39.8307	16.1290
4	21.5780	40.3498	17.1729
5	25.3272	41.0283	18.5601
6	28.8176	42.7063	20.5886
7	31.8774	44.6379	22.9004
8	34.8676	46.5684	25.2898
9	37.7255	48.5652	27.6495
10	40.6136	50.2731	29.7891
11	43.1811	51.7465	31.4742
12	45.5966	53.0583	32.8710

(b) supplc cFIR model

horizon [h]	critical point 1	critical point 2	critical point 3
1	-12.6672	20.6373	-8.9477
2	3.6866	30.6035	5.2472
3	13.2943	34.7397	11.4181
4	18.3019	35.9590	14.9800
5	23.0563	37.2037	18.1995
6	27.1536	39.3316	21.3865
7	30.5186	41.6330	24.4759
8	33.6480	43.8192	27.3243
9	36.5062	45.8661	29.8913
10	39.2566	47.4513	32.0293
11	41.7219	48.7801	33.5817
12	44.0663	50.1574	34.7064

been carried out by considering the case of the Roskilde district heating system over a period covering several years.

It has been explained that in contrast to the state-of-the-art statistical and physical approaches, the proposed model can account for varying time delays in the distribution network. Indeed, the described cFIR model belongs to the family of FIR models, but for which the model coefficients are replaced by smooth coefficient functions. The choice of the flux at the supply point as a variable that conditions the transfer function of the network is obvious. In contrast, the way to account for the social behavior of the consumers is not straightforward. This is why it has been chosen to evaluate if this should be accounted for within the cFIR model, i.e. by considering the time of the day as a second variable conditioning the cFIR model, or outside of the cFIR model just by allowing the offset term to have diurnal variations. The former alternative is computationally less expensive (owing



to the lower number of coefficients to be estimated) and actually had a significantly higher forecast accuracy for 2 out of the 3 critical points considered. Whatever the chosen approach, the prediction accuracy of our methodology is dramatically higher than that of the state of the art in statistical approaches to the present problem.

The proposed class of models may be used for simulation or prediction for other types of problems that involve nonlinearly varying time-delays in a transfer function. This could be the case in e.g. signal processing, network modeling, or flood modeling. In addition, the described method for the estimation of the coefficient functions in cFIR models can be improved in the future, both for offline and online applications. Mainly, the issue of optimally tuning the regularization parameter at each fitting point should be further investigated. This may involve the use of numerical methods, but one may also envisage to search for some analytical solutions similar to that existing for the choice of optimal regularization parameters for least squares estimators in linear models.

In parallel, it would be interesting to study the alternative possibility of considering cFIR models with the load (instead of the flux) variable as that which conditions the transfer function of the district heating system. The main interest of such an approach would then be that forecasts of the temperature at critical points could not be affected by the interdependence of the flux and temperature variables at the supply point. Broader perspectives to this work finally include the use of the described forecasting methodology within the model predictive control based methods employed today for decision-making in district heating systems. This will allow us to quantify the operational benefits of a higher accuracy of temperature forecasts.

## Acknowledgments

The methods and results gathered in the present report have been generated as part of the Center for Model-Based Control (CMBC) project, partly supported by the Danish Agency for Science Technology and Innovation (under the contract number 07-000795), which is hereby greatly acknowledged. The authors also gratefully acknowledge Roskilde Forsynings, company supplying district heating, electricity and power to the citizens of Roskilde Kommune, for providing the data used as input.

## References

- Arvatson, L. (2001). Stochastic modelling and operational optimization in district heating systems. Ph.D. dissertation, Lund Institute of Technology, Mathematical Statistics.
- Bishop, C.M. (1995). Training with noise is equivalent to Tikhonov regularization. *Neural Computation* **7**: 108-116.
- Cleveland, W.S., Devlin, S.J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association* **83**: 596-610.
- Dotzauer, E. (2002). Simple model for prediction of loads in district-heating systems. *Applied Energy* **73**:277-284.
- Golub, G.H., Hansen, P.C., O'Leary, D.P. (2000). Tikhonov regularization and total least squares. *SIAM Journal of Matrix Analysis and Applications*, **21**:185-194.
- Hastie, T., Tibshirani, R. (1990). *Generalized additive models*. Chapman & Hall/CRC: London.

- Hubing, N.E., Alexander, S.T. (1991). Statistical analysis of initialization methods for RLS adaptive filters. *IEEE Transactions on Signal Processing* **39**:1793-1804.
- Ismail, M.Y., and Principe, J.C. (1997). Equivalence between RLS algorithms and the Ridge regression technique. in: Proc. 13<sup>th</sup> Asilomar Conf. Signals, Systems and Computers (IEEE), pp. 1083-1087.
- Johansen, T.A. (1997). On Tikhonov regularization, bias and variance in nonlinear system identification. *Automatica* **33**:441-446.
- Leung, C.S., Young, G.H., Sum J., Wing-kay, K. (1999). On the regularization of forgetting recursive least squares. *IEEE Transactions on Neural Networks* **10**:1482-1486.
- Madsen H. (2006). *Time Series Analysis, 2nd Ed.*. Technical University of Denmark: Kgs. Lyngby. ISBN: 87-643-0098-6
- Nielsen, H.Aa. (2000). Parametric and non-parametric system modelling. Ph.D. dissertation, Technical University of Denmark, Informatics and Mathematical Modelling.
- Nielsen, H.Aa., Madsen, H. (2000). Predicting the heat consumption in district heating systems using meteorological forecasts. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, ENS.J.Nr. 1323/98-025.
- Nielsen, H.Aa., Nielsen, T.S., Joensen, A.K., Madsen, H., Holst, J. (2000). Tracking time-varying coefficient functions. *International Journal of Adaptive Control and Signal Processing* **14**:813-828.
- Nielsen, T.S. (2002). Online prediction and control in nonlinear stochastic systems. Ph.D. dissertation, Technical University of Denmark, Informatics and Mathematical Modelling, 2002.
- Sandou, G., Font, S., Tebbani, S., Hiret, A., Mondon, C. (2004). Global modelling and simulation of a district heating network. in: Proc. of the 8th International Symposium on District Heating and Cooling, Espoo, Finland.
- Sandou, G., Font, S., Tebbani, S., Hiret, A., Mondon, C. (2005). Predictive control of a complex district heating system. in: Proc. of the 44th IEEE Conference on Decision and Control, Seville, Spain.
- Sjöberg, J., Ljung, L. (1995). Overtraining, regularization and searching for a minimum, with application to neural networks. *International Journal of Control* **62**:1391-1407.
- Stone, M. (1974). Cross-validation and assessment of statistical predictions (with discussion). *Journal of the Royal Statistical Society B* **36**:111-147.
- Søgaard, H.T. (1993). Stochastic systems with embedded parameter variations - Applications to district heating. Ph.D. dissertation, Technical University of Denmark, Institute of Mathematical Statistics and Operations Research.
- Tikhonov, A.N., Arsenin, V.Y. (1977). *Solutions of Ill-posed Problems*. Wiscon: Washington, DC.
- Wang S.-G., Chow S.-C. (1989). A note on adaptive generalized ridge regression estimator. *Statistics & Probability Letters* **10**:17-21.

## A Detailed results related to multi-step ahead predictions with cFIR models and flux measurements as input

### A.1 Critical point 1

**TABLE 5:** Some prediction accuracy measures related to temperature forecasting at critical point 1, when the cFIR models are fed with flux measurements. These measures are the bias, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The first three error measures are expressed in °C while the latter one is in percentage of predicted output.

(a) rigid cFIR model

horizon [h]	bias [ $10^{-3}$ .°C]	MAE [°C]	RMSE [°C]	MAPE [%]
1	3.345	0.2555	0.3865	0.3535
2	4.126	0.2722	0.4147	0.3765
3	4.345	0.2832	0.4311	0.3918
4	4.158	0.2896	0.4410	0.4007
5	3.782	0.2940	0.4475	0.4068
6	3.126	0.2968	0.4512	0.4107
7	2.313	0.2987	0.4541	0.4133
8	1.469	0.3001	0.4559	0.4153
9	0.813	0.3009	0.4568	0.4163
10	0.313	0.3013	0.4574	0.4169
11	0.188	0.3015	0.4580	0.4171
12	0.094	0.3012	0.4576	0.4167

(b) supplc cFIR model

horizon [h]	bias [ $10^{-3}$ .°C]	MAE [°C]	RMSE [°C]	MAPE [%]
1	11.53	0.2773	0.4189	0.3842
2	17.82	0.3023	0.4541	0.4186
3	17.82	0.3081	0.4622	0.4265
4	17.13	0.3108	0.4661	0.4302
5	16.25	0.3127	0.4688	0.4327
6	15.10	0.3140	0.4702	0.4346
7	13.91	0.3152	0.4719	0.4362
8	12.78	0.3160	0.4732	0.4374
9	11.94	0.3167	0.4741	0.4383
10	11.28	0.3173	0.4750	0.4392
11	11.00	0.3178	0.4759	0.4399
12	10.88	0.3183	0.4765	0.4405

## A.2 Critical point 2

**TABLE 6:** Some prediction accuracy measures related to temperature forecasting at critical point 1, when the cFIR models are fed with flux measurements. These measures are the bias, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The first three error measures are expressed in °C while the latter one is in percentage of predicted output.

(a) rigid cFIR model

horizon [h]	bias [ $10^{-3}$ .°C]	MAE [°C]	RMSE [°C]	MAPE [%]
1	-2.31	0.3003	0.4330	0.4251
2	-2.74	0.3120	0.4535	0.4421
3	-3.37	0.3203	0.4659	0.4538
4	-4.14	0.3236	0.4719	0.4585
5	-5.00	0.3261	0.4763	0.4621
6	-5.74	0.3276	0.4796	0.4643
7	-6.43	0.3286	0.4819	0.4658
8	-7.03	0.3295	0.4836	0.4671
9	-7.60	0.3300	0.4846	0.4677
10	-8.03	0.3304	0.4851	0.4683
11	-8.26	0.3305	0.4852	0.4684
12	-8.26	0.3303	0.4851	0.4681

(b) supple cFIR model

horizon [h]	bias [ $10^{-3}$ .°C]	MAE [°C]	RMSE [°C]	MAPE [%]
1	15.86	0.3224	0.4707	0.4573
2	21.18	0.3431	0.5034	0.4866
3	20.75	0.3485	0.5114	0.4942
4	19.46	0.3510	0.5157	0.4979
5	18.01	0.3527	0.5185	0.5003
6	16.66	0.3540	0.5203	0.5021
7	15.52	0.3549	0.5215	0.5033
8	14.46	0.3554	0.5221	0.5040
9	13.86	0.3555	0.5225	0.5042
10	13.29	0.3557	0.5229	0.5045
11	13.06	0.3561	0.5235	0.5050
12	13.15	0.3566	0.5242	0.5056

### A.3 Critical point 3

**TABLE 7:** Some prediction accuracy measures related to temperature forecasting at critical point 1, when the cFIR models are fed with flux measurements. These measures are the bias, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The first three error measures are expressed in °C while the latter one is in percentage of predicted output.

(a) rigid cFIR model

horizon [h]	bias [ $10^{-3}$ .°C]	MAE [°C]	RMSE [°C]	MAPE [%]
1	-7.72	0.3209	0.4519	0.4414
2	-7.02	0.3526	0.4994	0.4862
3	-5.46	0.3825	0.5417	0.5276
4	-3.97	0.4017	0.5714	0.5547
5	-2.93	0.4170	0.5945	0.5761
6	-2.44	0.4269	0.6105	0.5899
7	-2.38	0.4335	0.6221	0.5991
8	-2.72	0.4379	0.6300	0.6052
9	-3.21	0.4409	0.6353	0.6094
10	-3.66	0.4425	0.6379	0.6116
11	-3.91	0.4427	0.6383	0.6120
12	-3.99	0.4418	0.6368	0.6107

(b) supple cFIR model

horizon [h]	bias [ $10^{-3}$ .°C]	MAE [°C]	RMSE [°C]	MAPE [%]
1	-28.03	0.3653	0.5187	0.5037
2	-17.95	0.3891	0.5536	0.5368
3	-16.88	0.4041	0.5745	0.5576
4	-15.42	0.4137	0.5885	0.5709
5	-13.71	0.4209	0.5991	0.5809
6	-11.94	0.4259	0.6069	0.5879
7	-10.32	0.4295	0.6128	0.5928
8	-9.13	0.4322	0.6172	0.5965
9	-8.40	0.4341	0.6206	0.5992
10	-8.18	0.4353	0.6230	0.6008
11	-8.21	0.4362	0.6247	0.6021
12	-8.18	0.4368	0.6257	0.6030

## B Detailed results related to multi-step ahead predictions with cFIR models and flux predictions as input

### B.1 Critical point 1

**TABLE 8:** Some prediction accuracy measures related to temperature forecasting at critical point 1, when the cFIR models are fed with flux predictions. These measures are the bias, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The first three error measures are expressed in °C while the latter one is in percentage of predicted output.

(a) rigid cFIR model

horizon [h]	bias [ $10^{-3}$ .°C]	MAE [°C]	RMSE [°C]	MAPE [%]
1	3.345	0.2555	0.3865	0.3535
2	2.813	0.2694	0.4077	0.3724
3	2.720	0.2893	0.4341	0.3998
4	3.564	0.3071	0.4572	0.4243
5	4.220	0.3197	0.4735	0.4415
6	4.251	0.3280	0.4834	0.4529
7	3.908	0.3337	0.4913	0.4607
8	3.470	0.3376	0.4967	0.4662
9	3.220	0.3401	0.5005	0.4696
10	3.407	0.3420	0.5033	0.4722
11	4.220	0.3436	0.5062	0.4743
12	4.595	0.3448	0.5084	0.4759

(b) supple cFIR model

horizon [h]	bias [ $10^{-3}$ .°C]	MAE [°C]	RMSE [°C]	MAPE [%]
1	11.38	0.2728	0.4127	0.3778
2	15.52	0.2931	0.4389	0.4055
3	15.36	0.3084	0.4585	0.4265
4	15.07	0.3222	0.4763	0.4453
5	14.60	0.3314	0.4879	0.4579
6	14.00	0.3375	0.4947	0.4663
7	13.21	0.3421	0.5011	0.4726
8	12.39	0.3454	0.5060	0.4771
9	11.76	0.3479	0.5103	0.4806
10	11.72	0.3499	0.5148	0.4834
11	12.13	0.3519	0.5192	0.4861
12	12.32	0.3537	0.5227	0.4887



## B.2 Critical point 2

**TABLE 9:** Some prediction accuracy measures related to temperature forecasting at critical point 1, when the cFIR models are fed with flux predictions. These measures are the bias, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The first three error measures are expressed in °C while the latter one is in percentage of predicted output.

(a) rigid cFIR model

horizon [h]	bias [ $10^{-3}$ .°C]	MAE [°C]	RMSE [°C]	MAPE [%]
1	-2.31	0.3003	0.4330	0.4251
2	-7.72	0.3057	0.4430	0.4322
3	-10.32	0.3229	0.4692	0.4564
4	-11.20	0.3395	0.4945	0.4797
5	-10.43	0.3518	0.5150	0.4971
6	-9.83	0.3588	0.5263	0.5070
7	-9.86	0.3635	0.5343	0.5137
8	-10.15	0.3671	0.5403	0.5187
9	-10.46	0.3684	0.5431	0.5206
10	-10.32	0.3702	0.5463	0.5231
11	-9.97	0.3719	0.5498	0.5255
12	-10.12	0.3732	0.5518	0.5273

(b) supple cFIR model

horizon [h]	bias [ $10^{-3}$ .°C]	MAE [°C]	RMSE [°C]	MAPE [%]
1	15.86	0.3224	0.4707	0.4573
2	18.38	0.3331	0.4864	0.4720
3	15.39	0.3491	0.5089	0.4946
4	13.20	0.3640	0.5309	0.5154
5	12.80	0.3744	0.5484	0.5300
6	12.52	0.3805	0.5573	0.5387
7	11.66	0.3842	0.5633	0.5440
8	10.80	0.3870	0.5681	0.5478
9	10.32	0.3883	0.5716	0.5497
10	10.49	0.3901	0.5773	0.5523
11	10.63	0.3925	0.5836	0.5557
12	10.35	0.3942	0.5859	0.5581

### B.3 Critical point 3

**TABLE 10:** Some prediction accuracy measures related to temperature forecasting at critical point 1, when the cFIR models are fed with flux predictions. These measures are the bias, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The first three error measures are expressed in °C while the latter one is in percentage of predicted output.

(a) rigid cFIR model

horizon [h]	bias [ $10^{-3}$ .°C]	MAE [°C]	RMSE [°C]	MAPE [%]
1	-7.72	0.3209	0.4519	0.4414
2	-7.42	0.3406	0.4811	0.4688
3	-7.54	0.3712	0.5252	0.5111
4	-5.46	0.3928	0.5590	0.5412
5	-3.02	0.4112	0.5871	0.5666
6	-0.67	0.4229	0.6071	0.5828
7	0.24	0.4312	0.6215	0.5944
8	0.89	0.4369	0.6316	0.6022
9	0.82	0.4405	0.6390	0.6073
10	0.64	0.4425	0.6425	0.6099
11	0.46	0.4431	0.6438	0.6109
12	0.09	0.4431	0.6437	0.6108

(b) supple cFIR model

horizon [h]	bias [ $10^{-3}$ .°C]	MAE [°C]	RMSE [°C]	MAPE [%]
1	-28.03	0.3653	0.5187	0.5037
2	-24.64	0.3732	0.5309	0.5140
3	-25.40	0.3903	0.5547	0.5375
4	-23.29	0.4030	0.5738	0.5550
5	-20.42	0.4132	0.5897	0.5690
6	-17.28	0.4200	0.6010	0.5783
7	-15.30	0.4249	0.6088	0.5850
8	-13.56	0.4285	0.6144	0.5899
9	-12.58	0.4311	0.6192	0.5935
10	-12.15	0.4324	0.6220	0.5954
11	-11.94	0.4336	0.6240	0.5970
12	-11.78	0.4349	0.6261	0.5988

## C Detailed results related to multi-step ahead predictions with the lagged transfer function approach of the state of the art

### C.1 Critical point 1

**TABLE 11:** *Some prediction accuracy measures related to temperature forecasting at critical point 1, with the lagged transfer function approach of the state of the art. These measures are the bias, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The first three error measures are expressed in °C while the latter one is in percentage of predicted output.*

horizon [h]	bias [ $10^{-3}$ .°C]	MAE [°C]	RMSE [°C]	MAPE [%]
1	9.95	0.2509	0.3663	0.3437
2	16.59	0.3160	0.4557	0.4329
3	21.74	0.3760	0.5288	0.5147
4	26.01	0.4175	0.5830	0.5712
5	29.89	0.4562	0.6341	0.6236
6	33.78	0.4917	0.6791	0.6717
7	38.65	0.5253	0.7212	0.7173
8	44.30	0.5597	0.7626	0.7639
9	50.12	0.5939	0.8037	0.8102
10	57.07	0.6297	0.8475	0.8584
11	65.10	0.6653	0.8909	0.9065
12	73.38	0.6991	0.9345	0.9519

## C.2 Critical point 2

**TABLE 12:** Some prediction accuracy measures related to temperature forecasting at critical point 1, with the lagged transfer function approach of the state of the art. These measures are the bias, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The first three error measures are expressed in °C while the latter one is in percentage of predicted output.

horizon [h]	bias [ $10^{-3}$ .°C]	MAE [°C]	RMSE [°C]	MAPE [%]
1	17.69	0.4272	0.5931	0.5973
2	30.35	0.5210	0.7009	0.7298
3	40.07	0.5897	0.7798	0.8249
4	48.36	0.6337	0.8290	0.8854
5	55.16	0.6712	0.8733	0.9375
6	61.79	0.7095	0.9186	0.9912
7	68.96	0.7498	0.9651	1.0477
8	76.88	0.7902	1.0112	1.1046
9	85.94	0.8295	1.0559	1.1600
10	95.77	0.8677	1.0986	1.2135
11	105.09	0.9028	1.1394	1.2628
12	113.81	0.9325	1.1755	1.3042

## C.3 Critical point 3

**TABLE 13:** Some prediction accuracy measures related to temperature forecasting at critical point 1, with the lagged transfer function approach of the state of the art. These measures are the bias, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The first three error measures are expressed in °C while the latter one is in percentage of predicted output.

horizon [h]	bias [ $10^{-3}$ .°C]	MAE [°C]	RMSE [°C]	MAPE [%]
1	7.45	0.3355	0.4761	0.4586
2	10.23	0.3943	0.5603	0.5408
3	12.30	0.4408	0.6262	0.6046
4	13.43	0.4736	0.6749	0.6500
5	14.72	0.5042	0.7209	0.6922
6	16.21	0.5323	0.7645	0.7309
7	17.92	0.5602	0.8061	0.7694
8	20.52	0.5865	0.8454	0.8057
9	23.66	0.6118	0.8832	0.8407
10	25.77	0.6335	0.9151	0.8708
11	27.60	0.6500	0.9395	0.8936
12	29.07	0.6638	0.9589	0.9130