



# Data assimilation as a key step towards the implementation of an efficient management of dissolved oxygen in land-based aquaculture

Edouard Royer<sup>1,2</sup>  · Roberto Pastres<sup>1</sup>

Received: 14 January 2022 / Accepted: 18 November 2022  
© The Author(s) 2022

## Abstract

A data assimilation (DA) methodology, e.g. the continuous-discrete Kalman filter (CD-KF), was applied to the assimilation of dissolved oxygen data, in order to obtain a dynamic estimation of the oxygen demand in a land-based aquafarm. The CD-KF was implemented on a dynamic model, which included as state variables the concentration of dissolved oxygen (DO) and fish respiration rate: the latter was considered as a non-observable stochastic variable. The model was applied to a 1-month long set of observations collected at a raceway rainbow trout farm, including (1) hourly time series of water temperature and dissolved oxygen concentration in the raceway influent and effluent and (2) a daily time series of fish number and fish weight distribution. The results show that the assimilation of DO data led to a dynamic estimate of DO demand which showed changes in the daily mean and the daily pattern: these were related to changes in the feeding regime. Furthermore, the methodology provided accurate short-term predictions of the DO concentration also in the presence of short-term fluctuations, which would be very difficult to relate to external forcings in a mechanistic model. These findings indicate that DA could be effectively used to design and implement efficient and robust control systems for optimizing the oxygen supply, thus contributing to the implementation of Precision Fish Farming in land-based aquafarms.

**Keywords** Precision Fish Farming · Rainbow trout · Data assimilation · Oxygen management · Land-based aquaculture · Kalman Filter

---

Handling editor: Gavin Burnell

---

✉ Edouard Royer  
e.royer@planetek.it

<sup>1</sup> Department of Environmental Sciences, Informatics and Statistics, Università Ca' Foscari Venezia, Via Torino 155, 30170 Venice Mestre, Italy

<sup>2</sup> Planetek Italia, Via Massaua 12, 70132 Bari, Italy

## Introduction

The dynamics of dissolved oxygen (DO) concentration is of key interest in land-based fish farming systems, as the oxygen supply through oxygenation or aeration is critical for fish welfare and growth (Lawson, 1995; Colt & Orwicz, 1991). Fish respiration is thus one of the most important processes to take into account when designing a land-based system: according to (Colt and Maynard, 2019) the daily pattern of fish respiration rate should be considered, in order to prevent low DO concentration conditions. The detection and prediction of short-term changes of the oxygen demand is also a key component of an adaptive oxygen control system that modulates the oxygen supply in response to the current demand.

The Precision Fish Farming (PFF) approach (Fore et al, 2018) aims at supporting optimal management decision-taking by feeding predictive models with real-time data concerning environmental variables, e.g. DO concentration, and “animal variables”, e.g. fish biomass and size distribution. Royer et al. (2021) applied this approach to a trout farm case study in Northern Italy, showing that this methodology can improve both the economic and environmental sustainability.

Rainbow trout metabolism, in particular the standard metabolic rate, shows a daily dynamics which is affected by both circadian rhythm and feeding regime (Bolliet et al., 2001; Heydarnejad & Purser, 2009). Bolliet et al. (2004) highlighted the importance of adequately phasing the food distribution with internal rhythms associated with nutrient partitioning. Royer et al. (2021) showed that the daily pattern of oxygen concentration in a trout farm can be simulated using a model of DO dynamics: in this study, fish respiration was modelled using a sinusoidal function. However, the daily mean value, phase and amplitude may vary from day to day, due to changes in feeding regimes, fish health and fluctuations in environmental, non-controllable, variables, e.g. water temperature and concentration of suspended particles. Such short-term variability can hardly be explained and predicted using process-based models. Thus, combining these models with a data-driven approach could be a flexible, robust and computationally efficient solution.

Data assimilation (DA) methods combine model predictions and observations in order to improve the estimation of the state of a given system, thus allowing a more reliable forecast (Bocquet, 2014; Niu et al., 2014). This goal is achieved by assuming that the state variables of a process-based model are, in fact, stochastic ones: non-observable variables, such as the parameters of a process-based model, can also be included in the state vector, and, therefore, their value can change in time. These changes can be driven by the information gained any time a new observation is collected. DA was introduced in statistics and engineering (Kalman, 1960), and its popularity has increased in several fields, such as meteorology (Ide et al. 1999), oceanography (Beck, 1987; Pastres et al. 2003) and ecology (Niu et al., 2014; Luo et al., 2011). DA methods are not widely used in the aquaculture field as yet. In this context, the assimilation of real-time observations can markedly improve the robustness and short-term predictive skill of dynamic models, which are the key components for implementing efficient data quality control and management systems, based on the PFF principles. As such, DA-based forecast and control systems could quickly expand in aquaculture, thanks to the availability of low-cost sensors able to monitor both environmental and animal variables (Fore et al., 2018).

To our knowledge, dynamic models including DA algorithms were not previously applied to the management of dissolved oxygen supply. The Kalman filter (KF) (Gelb 1974) is one of the most popular DA methods, which allows real-time estimation of

state variables and parameters of a system based on observational knowledge. The main goals of this study were (1) to test the applicability of the KF method and provide a protocol for its implementation in a land-based trout farm and (2) to show that a DO model including a DA algorithm could provide reliable short-time forecast of fish oxygen demand, thus leading to improve both economic efficiency and fish welfare, as sudden drops of DO could be avoided. The model was tested on a rainbow trout farm, as part of the activities of the GAIN (Green Aquaculture Intensification in Europe) H2020 EU-funded project.

## Materials and methods

### Kalman filter

The Kalman filter (KF) (Gelb, 1974; Kalman, 1960) is a well-known recursive estimation technique that has been used for decades in many different fields, including ecological modelling (Luo et al. 2011; Niu et al. 2014; Dabrowski, 2018). This DA methodology was applied also to aquaculture for monitoring fish behaviour (Pinkiewicz et al. 2011; Mehammer et al. 2018) and forecasting fish growth (Marafioti et al., 2012). The objectives of this method are (i) to improve the prediction of observable output variables, based on the information provided by any new datum, and (ii) to improve the estimation of the remaining state variables, which may include key model parameters. Key to the method is the assumption that the state variables are stochastic: in a process-based model, their dynamic is driven by a set of governing differential equations, whereas the introduction of a DA algorithm allows one to account for deviations from their trajectories, due to processes which are not explicitly taken into account by the mechanistic component of the model.

The KF was originally designed for linear models, but it can also be applied to non-linear models, by linearizing the model equations. The algorithm can assimilate continuous or discrete data. For instance, Marafioti et al. (2012) implemented a continuous-discrete extended Kalman filter (CD-EKF) in order to deal with non-linearity of the model and discrete measurements of the state.

In this work, as observations are discrete and the model is a linear continuous-time model, the continuous-discrete Kalman filter (Lewis et al. 2016) was used:

$$\dot{X}(t) = F(t).X(t) + B(t).U(t) + \xi(t) \quad (1a)$$

$$Z_k = H_k.X_k + v(t) \quad (1b)$$

In Eq. 1a,  $X$  is the augmented state (that includes not only the state of the process-based model but also some parameters of this model that becomes, in the KF perspective, part of the state),  $F$  is the system dynamics matrix,  $B$  is the forcing matrix,  $U(t)$  is a vector of external forcings and  $\xi(t)$  is a random system noise, such that  $\xi(t) \sim N(0, Q)$ , with  $Q$  being the system noise covariance matrix. Equation 1b is the measurement equation in which  $Z_k$  is the vector of variables measured at time  $k$ ,  $H_k$  the state-to-measurement matrix, which project the state space onto the space of the output variables, and  $v(t) \sim N(0, R)$  is the measurement noise.

The “best” estimate of the system state between two measurements is given by Eq. 1a. Once a new measurement is collected at time “ $k$ ”, the state is corrected based on Eq. 2:

$$X_{k:k} = X_{k-1:k} + v_k \tag{2}$$

where  $X_{k:k-1}$  is the value of the augmented state estimated using Eq. 1a,  $v_k$  is the innovation term and  $X_{k:k}$  is the corrected augmented state vector. The innovation term  $v_k$  is a weighted difference between the measurement and the ‘‘a priori’’ estimate of the system state:

$$X_{k:k} = X_{k-1:k} + K * (Z_k - H.X_{k-1:k}) \tag{3}$$

The weight  $K$ , i.e. the ‘‘Kalman gain’’, takes into account the variances of both the system and measurement noise: its dynamic is given by the following matrix differential equation:

$$K_k = P_{k:k-1}.H_k^T.[H_k.P_{k:k-1}.H_k^T + R_k]^{-1} \tag{4}$$

where  $P$  is the covariance matrix of the state and  $R$  is the variance of the measurement noise. In order to apply Eq. 4, it is necessary to solve the covariance matrix propagation in Eq. 5, which read as:

$$\dot{P} = F(t).P(t) + P(t).F^T(t) + Q \tag{5}$$

where  $Q$  is the system noise covariance matrix. The covariance matrix  $P$  is also updated after the collection of every new measurement using:

$$P_{k:k} = [I - K_k H_k]P_{k:k-1} \tag{6}$$

where  $I$  is the identity matrix.

In summary, the CD-KF algorithm includes three steps (Fig. 1):

- 1) Eqs. 1a and 5 are used for estimating the state and its covariance between  $t = ‘‘k-1’’$  and  $t = ‘‘k’’$ .
- 2) After collecting a new datum at  $t = k$ , (i) the Kalman gain is computed, using Eq. 4. The augmented state vector and the covariance matrix are corrected using Eqs. 3 and 6, respectively.

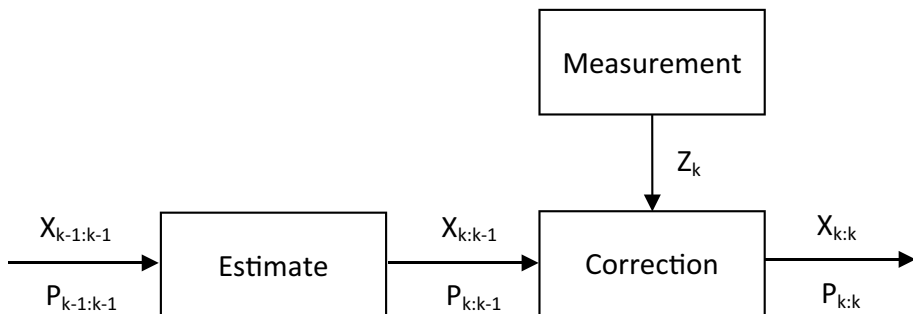


Fig. 1 The three steps of the Kalman filter methodology for data assimilation

### Application of CD-KF to the DO model

The KF methodology was applied for developing and testing a dynamic model for the prediction of DO concentration within the raceway and of oxygen demand due to fish respiration. The mechanistic model equations are presented in Table 1, which also explicitly shows the *F*, *H* and *K* matrices of Eqs. 1a and 1b. The DA model includes two state variables, namely: DO concentration,  $x_1$ , and the fish specific respiration rate at 15° C,  $x_2$ . The expected value of DO, Eq. 1 in Table 1, is estimated based on the following assumptions (see also Royer et al., 2021): (1) the raceway is well mixed; (2) fish respiration is the only significant contribution to oxygen consumption; (3) the rate of DO exchange with the atmosphere is proportional to the DO deficit (Ginot & Hervé, 1994). The expected value of the specific respiration rate is assumed to be constant (Table 1): its dynamics are driven entirely by the correction step of the KF algorithm. Water temperature is, in this case, a non-controllable model input: its effect on the respiration rate was modelled using an exponential function (Table 2) (Brigolin et al. 2010).

The matrix *H* and the explicit formulation of the Kalman gain, *K*, are given in Table 3, which also presents the parameters which specify the stochastic components of the model,

**Table 1** Process-based dynamic model simulating the evolution of dissolved oxygen and fish-specific respiration rate

State variables		
$x_1$	DO concentration in the raceway	[mg L <sup>-1</sup> ]
$x_2$	Fish specific respiration rate at $T_w=15$ °C	[mg kg <sup>-1</sup> h <sup>-1</sup> ]
State equation		
$\frac{dx_1}{dt} = Q(x_{in} - x_1)/V + S + k_{rear}(DO_{sat} - x_1) - MR_{15}exp(p_k(T_w - 15))/V$		
$\frac{dx_2}{dt} = \frac{dR_{15}}{dt} = 0$		
External forcings		
$T_w$	Water temperature	[°C]
$x_{in}$	Concentration of DO in the influent	[mg L <sup>-1</sup> ]
<i>M</i>	Fish biomass	[kg]
<i>Q</i>	Water flow rate	[L h <sup>-1</sup> ]
Vectors and matrices of the model		
$X(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$		
$F(t) = \begin{bmatrix} -(\frac{Q}{V} + k_{rear}) & -\frac{M}{V}exp(p_k(T_w - 15)) \\ 0 & 0 \end{bmatrix}$		
$U(t) = \begin{bmatrix} x_{in} \\ S \\ DO_{sat} \end{bmatrix}$		
$B(t) = \begin{bmatrix} \frac{Q}{V} & 1 & k_{rear} \\ 0 & 0 & 0 \end{bmatrix}$		
Control parameter		
<i>S</i>	Oxygen supply rate	[mg L <sup>-1</sup> h <sup>-1</sup> ]
Parameters		
<i>V</i>	Raceway volume	[L]
$k_{rear}$	Reaeration rate	[h <sup>-1</sup> ]

**Table 2** Functional expressions and parameters of the mechanistic component of the model

Functional expressions

$$R = R_{15} e^{\rho_k(T_w - 15)}$$

$$DO_{sat}(T_w) = 14.589 - 0.4T_w + 0.008T_w^2 - 0.0000661T_w^3$$

$$S = \frac{0.9 * LO2.P.MW_{O_2}}{N_a * k * [T_w + 273.15].V}$$

Parameters

$p_k$	0.07		Temperature coefficient	Myrick and Cech (2000)
$k_{rear}$	0.046	$h^{-1}$	Reaeration rate	Ciavatta et al. (2004)
$Q$	$1584 \times 10^3$	$Lh^{-1}$	Volumetric flow rate	
$LO2$	2400	$Lh^{-1}$	Oxygen supply rate	
$P$	101325	$Pa$	Atmospheric pressure	
$MW_{O_2}$	31.998	$gmol^{-1}$	Oxygen molar mass	
$N_a$	$6.022 \times 10^{23}$	$mol^{-1}$	Avogadro constant	
$k$	$1.38 \times 10^{-23}$	$JK^{-1}$	Boltzmann constant	
$V$	1280	$m^3$	Water volume in the raceway	
$f$	1/24	$Hz$		
$R_{15}$	This study	$mgh^{-1}kg^{-1}$	Oxygen consumption rate	
$DO_{sat}$		$mgL^{-1}$	Oxygen concentration at saturation	Genot and Hervé (1994)
$S$		$mgL^{-1}h^{-1}$	Oxygen concentration supply rate	

Initial values of the state variables

$$\begin{cases} x_1(t_0) \\ x_2(t_0) \end{cases} \text{with } \begin{cases} DO_{out}(t_0) = 8.82[mgL^{-1}] \\ R_{15}(t_0) = 101.5[mgh^{-1}kg^{-1}] \end{cases} \quad \text{Royer et al. (2021)}$$

**Table 3** Initial conditions for the covariance state matrix,  $P$ , and the system noise covariance one,  $Q$  (the latter is assumed to be constant in time), and state-to-measurement and Kalman gain matrices,  $H$  and  $K$

State covariance

$$P(t_0) = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix} \text{with } \begin{cases} \sqrt{\sigma_{11}} = 0.24[mgL^{-1}] \\ \sqrt{\sigma_{22}} = 100[mgh^{-1}kg^{-1}] \end{cases} \quad \text{Royer et al. (2021)}$$

System noise

$$Q(t) = Q = \begin{bmatrix} Q_{11} & 0 \\ 0 & Q_{22} \end{bmatrix} \text{with } \begin{cases} Q_{11} = 0.012[mg^2L^{-2}] \\ Q_{22} = 11.5[mg^2h^{-2}kg^{-2}] \end{cases} \quad \text{Pastres et al. (2003)}$$

$$R(t) = R \text{ with } \sqrt{R} = 0.2[mgL^{-1}]$$

State-to-measurement matrix and Kalman gain

$$H(t) = [1 \ 0]$$

$$K(t) = \frac{1}{\sigma_{11} + R} \begin{bmatrix} \sigma_{11} \\ \sigma_{12} \end{bmatrix}$$

also called “hyperparameters”. Their estimation in real world applications of the KF algorithm could be critical: in this paper, the methodology applied in Pastres et al. (2003) to a model simulating DO dynamics in a water body was used. The value of the parameter  $R_{15}$  was taken from Royer et al. (2021). The variances of the two state variables of the augmented state vector obtained from the calibration were subsequently used for estimating the system noise, represented by the diagonal matrix  $Q$ . The measurement noise,  $R$ , was estimated by averaging the absolute value of the differences between two consecutive measurements, in accordance with Woodall and Adams (1989).

### External forcings and system observations

The model described in the previous section was tested at a raceway trout farm located in Preore, (Trentino–Alto Adige, Italy): the, purposely designed, monitoring system is described in detail in Royer et al. (2021) and summarized below. The model requires as input time series of concentration of DO in the influent,  $x_{in}$ , water temperature,  $T_w$ , oxygen supply rate,  $S$ , and fish biomass,  $M$ . The latter is a key, time dependent, input, which connect the specific respiration rate,  $x_2$ , to the oxygen concentration,  $x_1$ : see first state equation in Table 1. In this application, the total biomass was calculated by multiplying the number of fish by the average fish weight. The former daily time series was obtained by subtracting from the initial number of fish the daily mortalities. The fish weight daily distribution was estimated using the non-invasive monitoring system, “Biomass Daily”, which is being used to estimate weight distribution in Atlantic salmon cages (Lopez Riveros, 2017) and was tested on rainbow trout during the H2020 project GAIN (Bolzonella et al, 2022).

The concentration of DO and water temperature were monitored at the inlet and outlet with an hourly frequency using two YSI EXO2 probes.

The model was applied to the data set collected from July 3, 2019, to July 31, 2019, in order to test its predictive capability under different feeding regimes. In fact, fish were not fed or underfed during the first 2 weeks, hereafter called “fasting weeks” due to the presence of a pathogen *Lactococcus garvieae* and regularly fed during the following 2 weeks, hereafter called “feeding weeks” with daily ration ranging from 0.4 (81 kg of feed) to 0.96% (200 kg of feed) of the average fish weight (see also Fig. 2 in the next section).

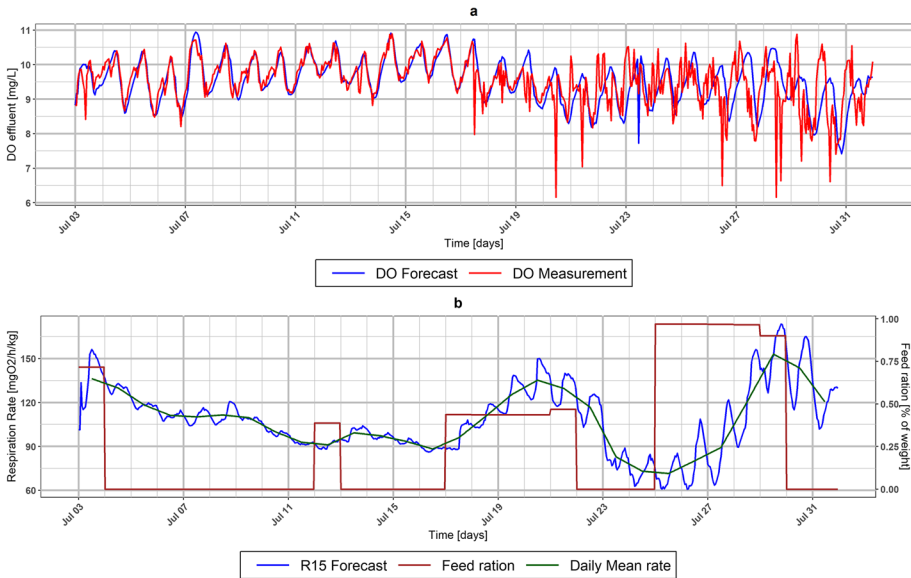
The model performance was assessed using the root mean square error, RMSE, as a goodness of fit indicator:

$$RMSE = \sqrt{\frac{\sum_{k=1}^{k=n} (x_{k-1:k} - y_k)^2}{n}} \tag{7}$$

where  $n$  is the number of observations,  $y_k$  is the DO concentration in the effluent at time  $t_k$ , and  $x_{k-1:k}$  is the DO concentration forecasted at time  $t_k$  using Eq. 1.

The time series of the innovation components ( $v_k$ ) were plotted and compared in terms of relative innovation, i.e. the ration between the innovation and the state estimate before the correction step, in order to assess the predictive skill of the mechanistic part of the model, and to better quantify the innovation itself.

The model and the CD-KF algorithm were coded using R software core (version 3.4.0) within R Studio (version 1.0.143). The differential equations were solved using the deSolve R package (version 1.21). The CD-KF matrix equations were entirely coded by the authors (none of the available KF packages for R was used).



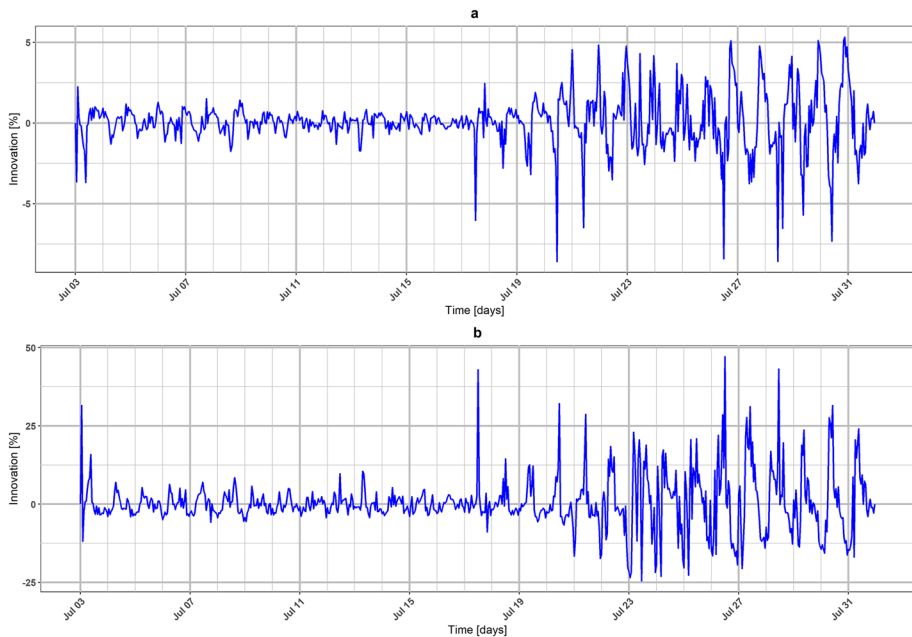
**Fig. 2** Evolution of the state variables: **a** predicted (blue line) and observed (red line) DO; **b** fish-specific respiration rate (blue line), its daily mean (green line) and, on the second axis, individual daily feed ration (brown line)

## Results

The results of the model application are summarized in Fig. 2, which shows the trajectories of the state variables, and in Fig. 3, which presents their relative innovations. A visual inspection of Fig. 2a, which presents the comparison between the DO estimated prior the correction step and the observations, indicates that the model succeeded in simulating the dynamics of the observations in both the “fasting” and “feeding” weeks. This is confirmed by the RMSE which, overall, was 0.27 [mgL<sup>-1</sup>], i.e. about 3% of the mean DO. The figure also shows that the daily pattern of DO observations was much more regular during the “fasting weeks”: in this time window, the model performed better (RMSE=0.12 mg/L). During the following “feeding weeks”, the daily range was, in general, higher, and isolated negative peaks were observed: the model, however, still provided reasonably accurate predictions, with an RMSE of 0.4 mg/L, i.e. about 5% of the mean DO.

Figure 2b shows the trajectory of the fish-specific respiration rate and, on the second axis, the daily feed ration, expressed as a percentage of the average fish weight. The respiration ranged from 20.12 to 208.45 [mg kg<sup>-1</sup> h<sup>-1</sup>], with an average value of 95.33 [mg kg<sup>-1</sup> h<sup>-1</sup>]. The fish oxygen demand is proportional to this non-observable variable, which showed both a clear daily pattern as well as changes in the mean level, which were entirely driven by the correction step and seems to be related mainly to the feeding regime. In fact, the daily range was markedly higher in the “feeding window”, compared with the “fasting” one. Furthermore, the mean level also increased soon after the feeding started and showed a sudden decrease when the farmer stopped the feeding again for 3 days, from July 22nd to 25th.





**Fig. 3** Time series of the relative innovations of DO **(a)** and fish-specific respiration rate, R15 **(b)**

This is confirmed by the analysis of the time series of the relative innovations, shown in Fig. 3. In fact, the relative innovation was much lower for both variables during the “fasting weeks”, compared to “feeding” ones, in which both innovations showed marked daily oscillations, perturbed by positive spikes: these are related to the irregular pattern of the DO observations and, in particular, to the isolated negative peaks shown in Fig. 2. The autocorrelation functions of both variables, not shown, presented a minimum at the lag of 12-h lag and a maximum at the lag of 24 h; this means that innovation cannot be considered as an uncorrelated “white” noise.

The mean values, standard deviations and ranges of the innovations and of the relative innovations, i.e. the ratio between the innovation  $\nu_k$  and the state estimate before the correction,  $X_{k-1:k}$ , are presented in Table 4. The standard deviation and range of the  $R_{15}$  relative innovation were significantly higher than those concerning DO, indicating that the introduction of the stochastic component leads to improve the model prediction and to estimate a relatively large fluctuation in the oxygen demand.

## Discussion

The findings presented in the previous section indicate that the data assimilation method tested in this paper allows one to:

- 1) obtain reliable predictions of dissolved oxygen concentration in a land-based system, during markedly different feeding regimes

**Table 4** Innovation statistics

	Statistics	[mg L <sup>-1</sup> ]	Relative [% of state estimate]
DO innovation	Mean	0.00	0.11
	SD	0.16	1.78
	Max	0.47	5.94
	Min	-0.97	-9.63
	Statistics	[mg h <sup>-1</sup> kg <sup>-1</sup> ]	Relative [% of state estimate]
R15 innovation	Mean	0.01	0.54
	SD	8.45	10.60
	Max	56.93	53.17
	Min	-21.56	-31.10

- 2) detect and obtain a quantitative estimate, in a real-world case study, of the oxygen demand, which results from the complex interplay of the circadian component of fish metabolism and the response to the feeding regime, often perturbed by other factors which could be hardly accounted for in a process-based model

The model proposed in this paper requires as an input the fish biomass density, which, in this case study, was estimated using a non-invasive monitoring system that may not be available in many farms. This may seem to limit the applicability of the methodology, but, in fact, the biomass is a smooth function, which could be interpolated from sampled data and/or estimated between samples using growth models (Bolzoni et al., 2022).

The CD-KF algorithm is a rather straightforward and efficient DA methods, which, however, is based on some hypothesis about the stochastic component of the model: in practical applications concerning ecology and environmental sciences, such hypothesis may be difficult to test or may not be fully complied with. In these cases, the KF is non-optimal, but this does not necessarily mean that its performance could be severely deteriorated: as the application of DA to aquafarming is still at a very early stage, in this paper, we tested the original KF formulation. The innovation statistics, see Table 4, indicates that, on average, the DA did not introduce a systematic bias in the state estimate. Furthermore, the standard deviation of the DO innovation was close to the estimated measurement noise, see Table 3, and below 2% of the model estimate, thus suggesting that, overall, the introduction of a stochastic component led to improve the predictive capability of the model, but the mechanistic component could still account for a large fraction of the data variability. The standard deviation of the innovation concerning the fish-specific respiration rate was higher than the DO one but, in relative terms, around 10% of the current system estimate and below 10% of its initial value. The mean respiration rate estimated for the whole period (July 3 to 31) was 95.33 [mg kg<sup>-1</sup> h<sup>-1</sup>], while Royer et al. (2021) estimated for the same period a mean of 101.5 [mg kg<sup>-1</sup> h<sup>-1</sup>]. The daily mean ranged from 53.82 to 137.29 [mg kg<sup>-1</sup> h<sup>-1</sup>], which is consistent with the findings presented in Alsop and Wood (1997) who measured a rate between 96 and 192 [mg kg<sup>-1</sup> h<sup>-1</sup>] and Eliason and Farrel (2014) who reported a rate of 75 [mg kg<sup>-1</sup> h<sup>-1</sup>]. Furthermore, the peak of 137.29 [mg L<sup>-1</sup>] was close to the value reported by Colt and Watten (1988).

The most severe violation of the basic hypothesis is given, in fact, by the autocorrelation of the innovation time series: in fact, the partial autocovariance function showed a daily

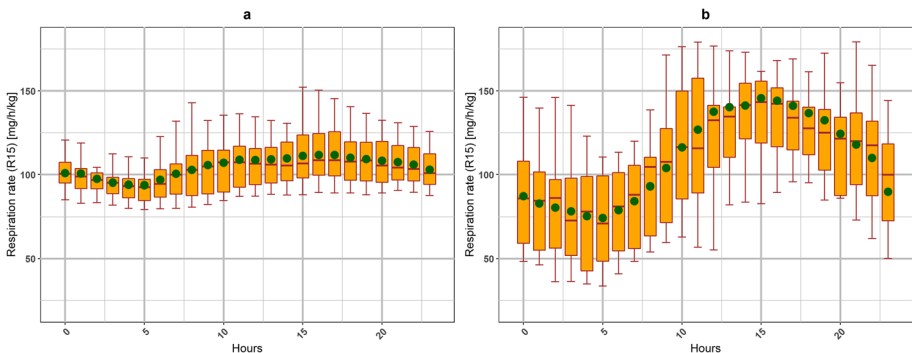
pattern, which, however, is already very evident from the time series of R15 in Fig. 2b. In principle, this points to the need of improving the mechanistic component of the model, by introducing a function which could mimic a systematic daily change of the specific respiration rate. Such a solution was suggested in Tudor (1999) and tested on the same data set used in this paper by Royer et al. (2021), who modelled this metabolic rate using the sinusoidal function:

$$R_{15} = R_m + A\cos[2\pi(t/24) + \varphi] \tag{8}$$

where  $t$  is the time [h],  $R_m$  is the average daily rate,  $A$  is the amplitude of the daily oscillation and  $\varphi$  is the phase. The implementation of Eq. 8 requires the estimation of three parameters which, however, can hardly be considered as constant in time, as they depend on the feeding regime. In fact, in Royer et al (2021),  $R_m$  was estimated using the data collected during the “fasting weeks”, but  $A$  had to be estimated for both the “fasting” and “feeding” weeks.

The comparison between the RMSE of the two models already suggests that the DA algorithm performs better than the process-based model: the CD-KF gave an RMSE of 0.12 vs 0.19 mg/L for the “fasting” weeks and of 0.44 vs 0.55 for the “feeding” ones. The improvement is more marked in the “feeding” weeks, which were characterized by higher daily oscillations of DO and higher noise: these were likely due to feeding anticipatory activity (Heydarnejad & Purser, 2009) and feed digestion. The DA model was able to track the effect of these processes and provided a dynamic estimate of the fish respiration rate which showed changes in both the daily mean and daily ranges.

As a result, the simulated DO followed the observations more closely, compared with the sinusoidal respiration model. The differences in the daily pattern of the fish respiration within the two time windows is clearly shown in Fig. 4, which presents its hourly distribution, estimated based on the results provided by the DA algorithm. The sinusoidal model (Eq. 8) performed quite well in the fasting weeks, characterized by a lower amplitude of the daily oscillation and lower variability. However, the same model, even after a retuning of two parameters, did not adequately simulate the daily respiration pattern under normal feeding. This could be due to both the choice of the function and/or the identifiability issues which arise in the simultaneous calibration of the two parameters,  $A$  and  $\varphi$ , and could lead to bias/uncertain estimates which are more noticeable as the noise increases.



**Fig. 4** Box plot summarizing the hourly distribution of the fish respiration rate: **a** for the “fasting weeks” time window; **b** for the “feeding weeks” one. The whole segment represents the range, the box indicates the interquartile range, and the horizontal bar is the median. Green bullets are the mean values

In fact, the retuning of the parameters led to estimate the same phase  $\varphi$  for the two time windows, thus failing in simulating the 6-h lag between the feeding time, about 9 a.m., and the respiration peak: this lag, however, is consistent with the literature (Colt & Watten, 1988). The retuning, indeed, led to estimate a higher absolute value of the amplitude  $A$  in Eq. 8 but did not allow to track the day-to-day variability of the hourly respiration rate which is shown in Fig. 4b.

The predictive capability of the CD-KF model could be used to design an advanced control system, based on the dynamic tuning of the oxygen supply, taking into account the residence time of the raceway, which was about 45 min. Therefore, a 1-h ahead prediction of the oxygen demand could already provide a useful input to such a system. Furthermore, the model output and, in particular, the innovation time series could provide relevant information for the quality control of the observations: such system could be very relevant for avoiding false alarms, in case sudden drops of DO concentration due to random fluctuation of the signal rather than to a real and potentially dangerous decrease due, for example, to a failure in the supply system. In practice, it may not be easy to identify random outliers by visual inspection or statistic on DO, and it could be easier to process the innovation, which, as can be seen in Fig. 3, shifts from highly positive to highly negative peaks when the observed DO (see Fig. 2a) suddenly drops but then goes back to acceptable levels. Since the parameter  $R_{15}$  has a physiological interpretation, as long as data are collected, it would be possible to establish a relative innovation threshold, for example, 2 times the standard deviation estimated on a representative data set, beyond which the observation is regarded as not reliable or suspicious.

## Conclusion

The implementation of aquafarm management practices based on the PFF approach requires the development of dynamic models for connecting the response of the “animal variables” to change in the environmental variables and control systems. Modelling approaches based on data assimilation perfectly fits this framework, as they allow one to update the estimates also of unobservable variables, based on real-time processing of the available observation.

To our knowledge, data assimilation was not previously applied to the estimation of fish respiration rate, a parameter that can hardly be measured in a real farming system. This task was accomplished using a basic version of the continuous-discrete Kalman filter, one of the simplest and most computationally efficient data assimilation method. Its application showed that it is possible to forecast the DO concentration in a trout raceway, as well as to obtain a dynamic estimate the fish respiration rate, using a model with a simple mechanistic component. The model was tested on a comprehensive data set, collected at a rainbow trout farm during a 1-month-long monitoring survey: during this period of time, the feeding regime drastically changed, from fasting to feeding the full ration recommended by the feeding tables. In spite of that, the model provided DO forecasts consistent with measurements as well as the mean and standard deviation of the respiration rate which were in line with the literature, without any need to retune the parameters. Furthermore, the CD-KF algorithm allowed to estimate changes in the mean daily respiration, correlated to the feed ration, as well as a consistent daily pattern, due to the interplay of the circadian rhythm of the fish metabolism and the time of feeding. The model results were also compared with those obtained using a more complex mechanistic model, in which the daily changes of the

fish respiration rate were simulated using a sinusoidal function, thus introducing two more but time-invariant parameters. As the model here presented performed better, we believe that the introduction of data-driven stochastic component in a simpler mechanistic model is a more efficient and robust solution to the problem of estimating fish oxygen demand and predicting DO concentration in a real-world land-based aquafarm.

The results presented in this paper suggests that this flexible modelling approach provides robust predictions of the respiration rate without the need of retuning the model in relations to changes in the feeding regime, in terms of quantity, i.e. daily ration, but also quality. This opens the way to the design and implementation of dynamic and efficient control system of the oxygen supply, which could be easily automatized. Such a system could contribute to the improvement of farm's sustainability: besides reducing the management costs, it may also lead to improve fish welfare and appetite, as sudden drops in the DO concentration could be avoided by matching the estimated demand with the supply.

**Acknowledgements** The authors would like to thank Matteo Leonardi for sharing the data presented in this paper and Joao G. Ferreira for his comments on the manuscript.

**Author contribution** All authors contributed to the data collection and to the study design. Coding activities as well as data formatting and analysis were performed by Edouard Royer. Supervision of the work was made by Roberto Pastres. The first draft of the manuscript was written by Edouard Royer, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** The research leading to these results has received funding from the European Union's H2020 Framework Programme for Research and Innovation, under Grant Agreement No. 773330.

**Data availability** The data sets analysed during the current study are available on the Zenodo repository at <https://doi.org/10.5281/zenodo.7309038>.

**Code availability** The code generated during the current study is described through the equations written in the manuscript.

## Declarations

**Ethics approval** Not applicable.

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alsop DH, Wood CM (1997) The interactive effects of feeding and exercise on oxygen consumption, swimming performance and protein usage in juvenile rainbow trout (*Oncorhynchus Mykiss*). *J Exp Biol* 200:2337–2346
- Beck MB (1987) Water quality modeling: a review of the analysis of uncertainty. *Water Resour Res* 23:1393–1442. <https://doi.org/10.1029/WR023i008p01393>

- Bocquet M (2014) Introduction to the principles and methods of data assimilation in the geosciences, 28 Jan 2021. <http://ceraa.enpc.fr/HomePages/bocquet/teaching/assim-mb-en.pdf>. Accessed 28 Jan 2021
- Bolliet V, Aranda A, Boujard T (2001) Demand-feeding rhythm in rainbow trout and European catfish synchronization by photoperiod and food availability. *Physiol Behav* 73:625–633. [https://doi.org/10.1016/S0031-9384\(01\)00505-4](https://doi.org/10.1016/S0031-9384(01)00505-4)
- Bolliet V, Jarry M, Boujard T (2004) Rhythmic pattern of growth and nutrient retention in response to feeding time in the rainbow trout. *J Fish Biol* 64:1616–1624. <https://doi.org/10.1111/j.0022-1112.2004.00418.x>
- Bolzonella M, Royer E, Lima A.C, Pastres R (2022) Predicting farmed rainbow trout weight distribution to improve feeding practice: an individual-based model approach. *Open Research Europe*. <https://doi.org/10.12688/openreseurope.14740.1>
- Brigolin D, Pastres R, Tomassetti P, Porrello S (2010) Modelling the biomass yield and the impact of seabream mariculture in the Adriatic and Tyrrhenian Seas (Italy). *Aquac Int* 18(2):149–163. <https://doi.org/10.1007/s10499-008-9232-4>
- Ciavatta S, Pastres R, Lin Z, Beck MB, Badetti C, Ferrari G (2004) Fault detection in a real-time monitoring network for water quality in the lagoon of Venice (Italy). *Water Sci Technol* 50(11):51–58
- Colt J, Maynard D (2019) Correction of metabolic parameters and unit process performance data Part I Derivation of equations. *J Aquac Eng Fish Res* 86:101999. <https://doi.org/10.1016/j.aquaeng.2019.101999>
- Colt J, Orwicz K (1991) Modeling production capacity of aquatic culture systems under freshwater conditions. *J Aquac Eng Fish Res* 10:1–29. [https://doi.org/10.1016/0144-8609\(91\)90008-8](https://doi.org/10.1016/0144-8609(91)90008-8)
- Colt J, Watten B (1988) Applications of pure oxygen in fish culture. *J Aquac Eng Fish Res* 7:397–441. [https://doi.org/10.1016/0144-8609\(88\)90003-9](https://doi.org/10.1016/0144-8609(88)90003-9)
- Eliason EJ, Farrell AP (2014) Effect of hypoxia on specific dynamic action and postprandial cardiovascular physiology in rainbow trout (*Oncorhynchus mykiss*). *Com Biochem Physiol, Part A* 171:44–50. <https://doi.org/10.1016/j.cbpa.2014.01.021>
- Fore M, Frank K, Norton T, Svendsen E, Alfredsen JO, Dempster T, Eguiraun H, Watson W, Stahl A, Sunde LM, Schellewald C, Skøien KR, Alver MO, Berckmans D (2018) Precision fish farming: a new framework to improve production in aquaculture. *Biosyst Eng* 173:176–193. <https://doi.org/10.1016/j.biosystemseng.2017.10.014>
- Gelb A (1974) *Applied Optimal Estimation*. MIT Press, Cambridge, Mass., p 374
- Genot V, Hervé JC (1994) Estimating the parameters of dissolved oxygen dynamics in shallow ponds. *Ecol Modell* 73:169–187
- Heydarnejad MS, Purser GJ (2009) Agonistics acts as a possible indicator of food anticipatory activity (FAA) in rainbow trout (*Oncorhynchus mykiss*). *Iran J Vet Res* 10:137–145. <https://doi.org/10.22099/ijvr.2009.1666>
- Ide K, Courtier P, Ghil M, Lorenc A (1999) Unified notation for data assimilation: operational, sequential and variational. *J Meteor Soc Japan* 75:181–189. [https://doi.org/10.2151/jmsj1965.75.1B\\_181](https://doi.org/10.2151/jmsj1965.75.1B_181)
- Kalman RE (1960) A new approach to linear filtering and prediction problems. *J Basic Eng* 82:35–45. <https://doi.org/10.1115/1.3662552>
- Lawson TB (1995) Oxygen and aeration. In: Lawson, T.B. (Ed.), *Fundamentals of aquaculture engineering*. Chapman & Hall, New York. 248–309. [https://doi.org/10.1007/978-1-4615-7047-9\\_11](https://doi.org/10.1007/978-1-4615-7047-9_11)
- Lewis FL, Lihua X, Dan P (2016) *Optimal and robust estimation*. CRC Press, Boca Raton, Florida, 552. <https://doi.org/10.1201/9781315221656>
- Lopez Riveros C (2017) Validation of high-accuracy of biomass estimator frames applied to the production of Atlantic salmon (biomass Daily), 5 Oct. 2020. <https://pentairaes.com/media/requestforquote/biomass-daily/docs/paes-biomass-validation-02.pdf>. Accessed 05 Oct 2020
- Luo Y, Ogle K, Tucker C, Fei S, Gao C, LaDeau S, Clark JS, Schimel DS (2011) Ecological forecasting and data assimilation in a data-rich era. *Ecol Appl* 21:1429–1442. <https://doi.org/10.1890/09-1275.1>
- Marafioti G, Alfredsen JA, Alver M (2012) Estimation of growth in farmed Atlantic salmon *Salmo Salar* L. based on dynamic energy budget (DEB) model and Kalman filtering. Poster at AQUA 2012 European Aquaculture Society (EAS) Conference 2012. [https://www.sintef.no/contentassets/231bb9f9d45047e09ff0bf5e03d32284/giancarlo\\_model.pdf](https://www.sintef.no/contentassets/231bb9f9d45047e09ff0bf5e03d32284/giancarlo_model.pdf). Accessed 05 Oct 2020
- Mehammer EB, Fore M, Sauder T, Chabaud VB, Parisini T (2018) Kalman estimation of position and velocity for ReaTHM testing applications. *J Phys Conf Ser* 1104:012008. <https://doi.org/10.1088/1742-6596/1104/1/012008>
- Myrick CA, Cech JJ (2000) Temperature influences on California rainbow trout physiological performance. *Fish Physiol Biochem* 22(3):245–254
- Niu S, Luo Y, Dietze MC, Keenan TF, Shi Z, Li J, Chapin FS III (2014) The role of data assimilation in predictive ecology. *Ecosphere* 5(5):65. <https://doi.org/10.1890/ES13-00273.1>

- Pastres R, Ciavatta S, Solidoro C (2003) The Extended Kalman Filter (EKF) as a tool for the assimilation of high frequency water quality data. *Ecol Modell* 170:227–235. [https://doi.org/10.1016/S0304-3800\(03\)00230-8](https://doi.org/10.1016/S0304-3800(03)00230-8)
- Pinkiewicz TH, Purserb GJ, Williams RN (2011) A computer vision system to analyse the swimming behaviour of farmed fish in commercial aquaculture facilities: a case study using cage-held Atlantic salmon. *Aquacult Eng* 45:20–27. <https://doi.org/10.1016/j.aquaeng.2011.05.002>
- Royer E, Faccenda F, Pastres R (2021) Estimating oxygen consumption of rainbow trout (*Oncorhynchus mykiss*) in a raceway: a Precision Fish Farming approach. *Aquacult Eng* 92:102141. <https://doi.org/10.1016/j.aquaeng.2020.102141>
- Tudor M (1999) Diurnal changes of dissolved oxygen in fouling land-based tanks for rearing of sea bass. *J Aquac Eng Fish Res* 19:243–258
- Woodall WH and Adams BM (1989) Statistical process control. In: Wadworth, Jr., H.M. (Ed.), *Handbook of Statistical Methods for Engineers and Scientist*. McGraw-Hill, New York

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.