A hybrid reinforced learning system to estimate resilience indicators

Simon Enjalbert, Frédéric Vanderhaegen

A hybrid reinforced learning system to estimate resilience indicators

Simon Enjalbert[a,b,c], Frédéric Vanderhaegen[a,b,c]

[a] *Univ Lille Nord de France*
[b] *UVHC, LAMIH*
[c] *CNRS UMR 8201, F-59313 Valenciennes*

Abstract

This paper describes a learning system based on resilience indicators. It proposes a hybrid learning system to estimate Human-Machine System performance when facing unprecedented situations. Collected data from various criteria are compared with data estimated using the local and the global resilience indicators, to give both instantaneous and over-time Human-Machine System states. The learning system can be composed of two different, separate reinforcement functions; the first allowing reinforcement of its own system knowledge and the second allowing reinforcement of its estimation function. When used together in a hybrid approach, the resilience indicator estimation should be improved. The learning system is then applied in a simulated air transport context and the impact of each reinforcement function on resilience indicator estimation is assessed. The hypothesis on performance of hybrid reinforcement learning is confirmed and it provides better results than those obtained by the knowledge based reinforcement or the estimation based reinforcement alone.

Keywords: Resilience engineering, Learning, Man-Machine systems.

Preprint submitted to Engineering Applications of Artificial Intelligence *July 20, 2017*

1. Introduction

Ouedraogo et al. (2013) defined resilience as the positive ability of a Human-Machine System (HMS) to recover from or adapt to critical situations. The recovery function consists of getting back to the previous normal functioning state and the adaptation function aims to provide the system with a new stable functioning state. A large amount of research has been performed in research laboratories about system safety and security in transport or industry based on this concept (Orwin and Wardle, 2004; Pérez-España and Arreguın-Sánchez, 2001; Enjalbert et al., 2013; Cacciabue et al., 2013). Some of this research involves assessment based on various criteria. These system evaluation criteria mainly concern human or machine behaviours or their effects, or the occurrence or consequences of external perturbations. These effects or perturbations relate for instance to human workload (Vanderhaegen, 1997), to human errors (Lin et al., 2015), to the quality or the production of services (Polet et al., 2009), and to the quality of cooperation or learning activities (Vanderhaegen et al., 2006). Therefore, resilience emerges in a risk management process and relates to the system capacity to survive both planned and unexpected hazardous events (Enjalbert et al., 2011). Unprecedented situations are defined as events with a very low frequency of occurrence and/or which may have catastrophic consequences for HMS.

This paper focuses on the learning system developed to estimate resilience indicators. The reinforcement functions of the learning system concern reinforcement of the system knowledge and reinforcement of the estimation parameters. This hybrid approach has been developed and tested on a flight simulator during an in-flight refuelling activity involving a team composed of four people. Several unexpected events with potential catastrophic consequences are incorporated and data collected on HMS during unprecedented situations are used by the hybrid reinforced learning system to estimate the local and the global resilience indicators.

In the second section of the paper, the need for resilience indicators and the principles of the reinforced iterative learning approaches are presented in order to introduce the contribution of the present work. In the third section, the generic architecture of the learning system is detailed with specific focus on reinforcement functions. Finally, a validation example showing the impact of reinforcement functions on resilience indicator estimation is described and the effectiveness of hybrid reinforcement is demonstrated.

2. Learning approaches and resilience assessment

Several concepts of learning can be found in literature. For instance, learning by imitation or observation consists in copying a given behaviour, or a sequence or a repetition of behaviours (Chella et al., 2006; Calinon et al., 2007). When facing a new situation for which no knowledge is defined, trial-and-error based learning should be applied (Rose et al., 2014). A redundant learning system is another way to engage the learning capacity of the system (Vanderhaegen and Zieba, 2014). Cooperative learning or co-learning are then useful for exchanging data between decision makers in order to understand the learning process or to share knowledge (Doisy et al., 2014). Effective techniques, characterized by efficient self-learning and adaptivity abilities, have been employed to construct learning systems (Xu and

Yan, 2004; Liu et al., 2013; Norrlöf and Gunnarsson, 2005; Wiering and van Hasselt, 2008). Many of these involve reinforcement learning or reinforced learning. Reinforcement learning is usually applied for repetitive tasks, in order to minimize tracking errors. If the error reduction is successful, the reinforcement is based on a reward for managing knowledge. Other authors prefer using the vocabulary of reinforced learning because their interest is not limited to repetitive tasks and error tracking reduction. Vanderhaegen et al. (2011) focused on the learning from human errors in order to provide human operators with decision support tools .

In this study, the learning approach objective is to estimate missing or immeasurable data from Human-Machine System facing unprecedented situations. In the first section, a theoretical analysis based on extended State of the Art of Iterative Learning Control (ILC) systems is proposed. Then, in a second section, indicators based on resilience criteria for HMS are developed. Finally, these indicators are adapted to reinforced learning approaches.

2.1. Iterative learning control systems

The feedforward process aims at assessing the future possible decisions regarding the current system states and the management of the previous ones. The feedback aims at recovering possible erroneous knowledge, at refining knowledge or at creating new knowledge (Vanderhaegen, 2010). So the feedforward-feedback mechanism that consists in using the current knowledge related to previous activities in order to calculate the future ones. A great number of research works have proposed feedback and/or feedforward controllers using different methods in order to reach the mentioned objectives. There are frequency based approach (related to iteration frequency) or temporal based approach (related to timing process).

Iterative Learning Control (ILC) systems are used to benefit from the repetitive nature of the tasks as experience gained to compensate for the poor or incomplete knowledge of the plant model and disturbance. The repeatability of the task determines the learning ability of the ILC. Current (e_i) in Equation (1) and previous (e_{i-1}) in Equation (2) tracking errors , and previous input u_{i-1} are used to assess the current input u_i in Table 1. The recursive process of ILC technique to assess the current characteristics and to improve tracking control performance in batch processes is given in Equation (3). The formalism can be seen as a generalization of the previous ones; the control is done regarding the previous errors at certain level because of limited memory capacity. A feedback-feedforward structure for the trajectory tracking of a linear Direct Current motor is given in Equation (4). The same structure for sharp tracking control of a manipulator robot, by employing a saturated input γ which limits the control input within a reasonable bound, was also proposed. The corresponding learning control updating law is given by Equation (5). The class of non-linear systems to which the proposed learning scheme can be applied is then extended. A combined feedback-feedforward controller and disturbance observer designed for a direct drive motion control was proposed in Equation (6). The digital disturbance observer is included in the proposed feedbackfeedforward control structure to compensate for disturbances (friction and cogging effects). Finally, a framework for the assessment of the consequences of human errors based on learning and prediction of the actions of a human operator is given in Equation (7) in Table 1. These processes are modelled by using the itera-

Table 1: Different formalisms for feedforward and/or feedback based learning control.

References	Formula
Xu et al. (2004) Ojha et al. (2017) Geng et al. (2017)	$$u_i = u_{i-1} + G_{feedforward}(e_{i-1}) \quad (1)$$
Xu et al. (2004) Radac and Precup (2016)	$$u_i = u_{i-1} + G_{feedback}(e_i) \quad (2)$$
Lee and Lee (2007)	$$u_i = u_{i-1} + G_1(e_{i-1}) + \cdots + G_p(e_{i-p}) \quad (3)$$
Lee et al. (2000)	$$u_i = u_{i-1} + G_{ff}e_{i-1} + G_{fb}e_i \quad (4)$$
Jang et al. (1995)	$$u_i = \gamma\nu_i = \gamma(u_{i-1} + G_{ff}e_{i-1} + G_{fb}e_i) \quad (5)$$
Yan and Shiu (2008)	$$\begin{aligned} u_i &= u_i^{ff} + u_i^{fb} - u_i^{d} \\ &= G_{ff}(e_{i-1}, u_{i-1}) + G_{fb}e_i - G_d(e_{i-1}, u_{i-1}) \end{aligned} \quad (6)$$
Vanderhaegen et al. (2009) Polet et al. (2012)	$$u_i = e_i + G((e_{i-1}, u_{i-1}), \ldots, (e_0, u_0)) \quad (7)$$

tive learning control concept and by integrating it in a feedforward-feedback approach.

ILC has become a competitive control method through the development of different learning controllers for many applications, essentially in robotic operations, chemical processes and motor drive machines. Initially the ILC input signal is formed using the error from previous iterations, *i.e.*, the input u_i is computed using the previous input u_{i-1} and e_{i-1} in so-called Previous Cycle Learning (PCL) in Equation (1) or recursively $e_{i-1}, \ldots, e_{i-p}$ in Equation (3). Several authors have computed the input u_i using the current tracking error e_i in so-called Current Cycle Learning (CCL) in Equation (2). Then, it has been proposed to combine the current error, e_i with the previous one e_{i-1}, when forming u_i in Equation (4), (5) and (6). This approach leads to a causal relationship between the current error and the input signal. It can be seen that PCL and CCL are functioning a complementary manner with the aim to improve the control performance through Previous and Current Cycle Learning (PCCL) structure, complementary role of feedback and feedforward structures.

The formalisms, summarised in Table 1, are used to deal with machines processes control (optimize robot or motor motion) during repetitive tasks mostly tracking errors performance control by managing a static knowledge. These control processes are not applied to problems involving humans and do not manage knowledge in unexpected or unprecedented situations. An extended approach by using the previous couples $((e_{i-1}, u_{i-1}), \ldots, (e_0, u_0))$ was proposed with feedforward-feedback learning control systems having their updating laws mostly depending on current and/or previous errors in Equation (7). The originality of this model is that it is applied to HMS with the aim to predict human errors. It combines feedforward-feedback processes and use predefined knowledge that is reinforced or corrected regarding the observed previous couples.

A State of the Art has been realized to compare different structures of the feedforward and/or feedback Iterative Learning Control systems in order to select the more appropriate one or to build an efficient one, for improving knowledge on known situations and for creating knowledge related to new situations. Therefore, the proposed article extends the Iterative Learning Control concept by proposing a hybrid reinforced learning structure that reinforces the learning by controlling two criteria of learning errors: errors between knowledge and error between predictions by taking into account matrices of data instead of vectors of data. Moreover, this new structure is applied to predict resilience indicators.

2.2. Resilience indicators

The proposed learning contribution should be able to estimate instantaneous and over-time HMS states, called respectively the local and the global resilience indicators. These indicators are based on several criteria such as the success level of a given task, the safety level of this task or the human workload in terms of interactions with the technical systems. For an iteration i and k criteria of resilience, the vector denoted U_{ki} in Equation (8) is based on two indicators, the local indicator, u_{ki}, and the global indicator, $\sum_{i=1}^{i} u_{ki}$.

$$U_{ki} = \begin{pmatrix} u_{ki} \\ \sum_{i=1}^{i} u_{ki} \end{pmatrix} \tag{8}$$

An iteration i is a discrete event at the end of a time window, that enables synchronization between HMS to be compared depending on the situation. The term U_i is then a vector with k rows, representing the k resilience criteria for iteration i in Equation (9).

$$U_i = \begin{pmatrix} U_{1i} \\ U_{2i} \\ \vdots \\ U_{ki} \end{pmatrix} \tag{9}$$

Finally, in Equation (10), the matrix U_I is incremented, by iteration 1 to i with i columns based on U_{ki} vectors.

$$U_I = \begin{pmatrix} U_{11} & U_{12} & \dots & U_{1i} \\ U_{21} & U_{22} & \dots & U_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ U_{k1} & U_{k2} & \dots & U_{ki} \end{pmatrix} \tag{10}$$

2.3. Reinforced learning systems

Reinforcement or reinforced learning systems are mainly based on the iterative learning control principle illustrated in Figure 1 with K the content of a knowledge composed of K_I matrices from precedent 'similar' situations.

Iterative learning control consists, for iteration i and for each of the k resilience criteria, of using the previous tracking errors ε_{ki-1} from a given reinforcement function $G_{feedback}$ and the previous state of HMS in terms of resilience indicators u_{ki-1} from U_{i-1} in U_{I-1} to estimate the current state of HMS u^*_{ki}. This process is repeated for the k resilience criteria and can be generalized as in Equation (11).

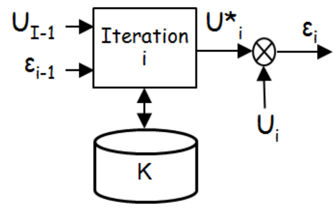

Figure 1: Estimation process in learning system architecture.

$$U_i^* = U_{i-1} + G_{feedback}(\varepsilon_{i-1}) \tag{11}$$

The tracking error, denoted ε_i for iteration i, is assessed in Equation (12) by evaluating the differences between the actual vector U_i from matrix U_I and the estimated vector U_i^* for the k resilience criteria.

$$\varepsilon_i = \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \vdots \\ \varepsilon_{ki} \end{pmatrix} = \begin{pmatrix} U_{1i}^* - U_{1i} \\ U_{2i}^* - U_{2i} \\ \vdots \\ U_{ki}^* - U_{ki} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} u_{1i}^* - u_{1i} \\ \sum_{i=1}^{i} (u_{1i}^* - u_{1i}) \end{pmatrix} \\ \begin{pmatrix} u_{2i}^* - u_{2i} \\ \sum_{i=1}^{i} (u_{2i}^* - u_{2i}) \end{pmatrix} \\ \vdots \\ \begin{pmatrix} u_{ki}^* - u_{ki} \\ \sum_{i=1}^{i} (u_{ki}^* - u_{ki}) \end{pmatrix} \end{pmatrix} \tag{12}$$

Other approaches integrate a feedforward-feedback process in order to take into account possible future tracking errors with another function called $G_{feedfoward}$ in Equation (13):

$$U_i^* = U_{i-1} + G_{feedforward}(\varepsilon_i) + G_{feedback}(\varepsilon_{i-1}) \tag{13}$$

A way to simplify this approach in Equation (14) considers the tracking errors as input data ε_i, and the current result of the function (output) is defined by combining the previous pairs $(\varepsilon_{i-1}, U_{i-1})$:

$$U_i^* = \varepsilon_i + G_{feedback}(\varepsilon_{i-1}, U_{i-1}) \tag{14}$$

The current and previous errors ε_i, ε_{i-1}, ... ε_1, considered as inputs for estimation of HMS resilience indicators, can be composed from the benefits, costs and potential deficits or dangers of a given human action (Vanderhaegen et al., 2011).

At a given iteration i, the reinforced learning system aims to estimate the local and global indicator values, termed U_i^* in Equation (15) relating to the previous actual change in U_{I-1}, based on the errors between the previous estimation, the actual value and the content of a knowledge K:

$$U_i^* = G_{feedforward}(U_{I-1}) + G_{feedback}(U_{I-1}, U_{i-1}^*, K) \tag{15}$$

The knowledge K is composed of K_I matrices, each containing the local and the global indicator values for iterations 1 to i. The structure of K_I is then the same as that of matrix U_I for iteration i.

Finally, two reinforcement processes are combined to form the iterative learning function: the first consisting of reinforcing the associated knowledge, and the second aiming to reinforce the error reduction between the estimated and actual values. Each reinforcement impact will be individually assessed, then both reinforcements will be combined in order to produce a so-called hybrid reinforcement function taking into account both local and global previous resilience indicators values *i.e.* a multi objective hybrid reinforcement as suggested by Delgado et al. (2008).

3. Principle of the hybrid learning system

3.1. Generic architecture of the learning system

The generic hybrid learning system architecture is an iterative process, depicted in Figure 2. Each iteration in the $G_{feedfoward}$ function aims to determine the presence of a similar matrix U_{I-1} in the knowledge K with the objective of estimating local and global resilience indicators U_i^*. In terms of the inputs U_{I-1}, the $G_{feedfoward}$ function searches for the K_{I-1} in K, denoted E_{I-1}, that is similar to U_{I-1}, in order to identify vector U_i^* that is assumed to be equal to E_i.

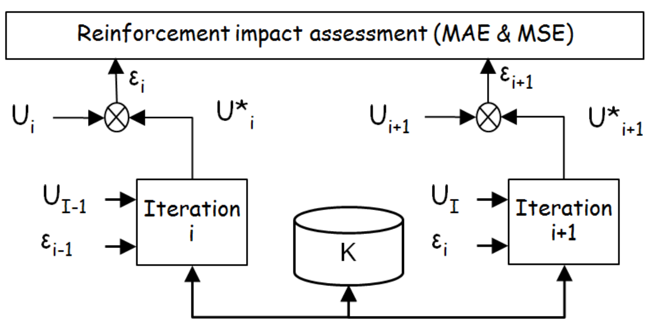

Figure 2: The generic hybrid learning system architecture.

The initial estimation of U_i^* is realized in Equation (16) by applying this $G_{feedfoward}(U_{I-1})$ function based on the Euclidean norm:

$$\begin{gathered} G_{feedforward}(U_{I-1}) = U_i^*, \\ U_i^* = E_i, \\ \forall K_{I-1} \in K, ||E_{I-1} - U_{I-1}|| = Min||K_{I-1} - U_{I-1}|| \end{gathered} \tag{16}$$

After this, two kinds of reinforcement could be applied to improve estimation performance by using the input matrix U_{I-1} and previous error ε_{i-1}:

- Reinforcement of the current knowledge.
- Reinforcement of the current estimation.

To quantify the impact of the reinforcement functions, the difference between the Mean Absolute Errors (MAE), and the Mean Squared Error (MSE) are assessed. The reinforcement processes aim to refine this estimation by integrating the results of the $G_{feedback}(U_{I-1}, U^*_{i-1}, K)$ function. At a given iteration i, the first reinforcement integrates the knowledge K by considering the errors between the inputs U_{I-1} and the knowledge content, whereas the second reinforcement process takes into account the ε_{i-1} error between U_{i-1} and U^*_{i-1} made at the previous iteration $i-1$. The corresponding $G_{feedback}$ functions are then respectively renamed G_{fbk_k} in Equation(17) for Knowledge reinforcement and G_{fbk_e} in Equation (18) for Estimation reinforcement, and are defined with the following inputs:

- Knowledge reinforcement:

$$G_{feedback}(U_{I-1}, U^*_{i-1}, K) = G_{fbk\_k}(U_{I-1}, K) \tag{17}$$

- Estimation reinforcement:

$$G_{feedback}(U_{I-1}, U^*_{i-1}, K) = G_{fbk\_e}(U_{I-1}, U^*_{i-1}) \tag{18}$$

3.2. Knowledge reinforcement

The first reinforcement principle aims to reinforce the knowledge content related to the gaps between the matrix content K_{I-1} and the inputs U_{I-1}. The reinforcement consists, for iteration i, of searching for the winner vector, denoted W_{i-1} in W_{I-1} in K_{I-1}, which is similar to the input vector U_{i-1} in U_{I-1}. The knowledge reinforcement then proceeds in two steps. The first step, G_{fbk_node}, aims to reinforce the W_{I-1} matrix content and the second step, G_{fbk_base}, consists of merging this new knowledge with the other knowledge matrices in Equation (19).

$$G_{fbk\_k}(U_{I-1}, K) = G_{fbk\_base}(G_{fbk\_node}(U_{I-1}, K)) \tag{19}$$

The $G_{fbk\_node}(U_{I-1}, K)$ function in Equation (20) aims to identify the W_{i-1} vector in W_{I-1} defined as follows:

$$\begin{gathered} G_{fbk\_node}(U_{I-1}, K) = W_{I-1}, \\ \forall K_{I-1} \in K, ||W_{I-1} - U_{I-1}|| = Min||K_{I-1} - U_{I-1}|| \end{gathered} \tag{20}$$

The reinforcement of W_{I-1} is achieved with the $G_{fbk\_node}(U_{I-1}, K)$ function in Equation (21):

$$\begin{aligned} G_{fbk\_node}(U_{I-1}, K) &= W_{I-1}^{reinforced}, \\ W_{I-1}^{reinforced} &= W_{I-1} + (W_{I-1} - U_{I-1}) \\ &= 2 * W_{I-1} - U_{I-1} \end{aligned} \tag{21}$$

The reinforcement of the entire knowledge is performed using the $G_{fbk\_base}(U_{I-1}, K)$ function in Equation (22):

$$
\begin{aligned}
G_{fbk\_base}(U_{I-1}, K) &= K_{I-1}^{reinforced}, \\
\forall K_{I-1} \in K, K_{I-1}^{reinforced} &\neq W_{I-1}^{reinforced}, \\
K_{I-1}^{reinforced} &= K_{I-1} + (K_{I-1} - W_{I-1}^{reinforced}) \\
&= 2 * K_{I-1} - W_{I-1}^{reinforced}
\end{aligned}
\tag{22}
$$

The estimation of U_i^* becomes, in Equation (23):

$$
\begin{aligned}
U_i^* &= G_{feedforward}(U_{I-1}) \\
&\quad + G_{fbk\_base}(G_{fbk\_node}(U_{I-1}, K))
\end{aligned}
\tag{23}
$$

3.3. Estimation reinforcement

For all iterations, the estimation error between U_i and U_i^*, denoted ε_i, is assessed using the $G_{fbk\_e}(U_I)$ function in Equation (24) that is dedicated to this vector assessment:

$$
G_{fbk\_e}(U_I, U_i^*) = \varepsilon_i \tag{24}
$$

If the error of the previous iteration noted ε_{i-1} is known, then, this assessment can be approximated in Equation (25) as follows:

$$
\begin{aligned}
G_{fbk\_e}(U_I, U_i^*) = \varepsilon_i &= \frac{1}{2} * (\varepsilon_i + \varepsilon_{i-1}) \\
&= \frac{1}{2} * (\varepsilon_{i-1} + U_{i-1} - U_{i-1}^*)
\end{aligned}
\tag{25}
$$

Therefore, the estimation of U_i^* is carried out using this error and by applying the $G_{feedforward}(U_{I-1})$ function in Equation (26):

$$
U_i^* = G_{feedforward}(U_{I-1}) + \varepsilon_i \tag{26}
$$

3.4. Hybrid reinforcement

The hybrid reinforcement integrates both the knowledge reinforcement and the estimation reinforcement. The $G_{feedback}(U_{I-1}, U^*_{i-1}, K)$ function is composed of the $G_{fbk\_e}(U_{I-1}, U^*_{i-1})$ function and the $G_{fbk\_k}(U_{I-1}, K)$ function, in Equation (27):

$$\begin{aligned} &G_{feedback}(U_{I-1}, U^*_{i-1}, K) \\ &= G_{fbk\_e}(U_{I-1}, U^*_{i-1}) + G_{fbk\_k}(U_{I-1}, K) \end{aligned} \tag{27}$$

After integrating the component functions characteristics, this function becomes, in Equation (28):

$$\begin{aligned} &G_{feedback}(U_{I-1}, U^*_{i-1}, K) \\ &= \varepsilon_i + G_{fbk\_base}(G_{fbk\_node}(U_{I-1}, K)) \end{aligned} \tag{28}$$

The global application for determining U^*_i is then given by Equation (29):

$$\begin{aligned} U^*_i &= G_{feedforward}(U_{I-1}) + \varepsilon_i \\ &+ G_{fbk\_base}(G_{fbk\_node}(U_{I-1}), K) \end{aligned} \tag{29}$$

4. Evaluation of hybrid reinforced learning system

The Hybrid reinforced learning architecture was tested using a military air transport system, involving a simulated cockpit with a four-person flight crew, illustrated in Figure 3. The experiments were performed with the in-flight refuelling group from the Istres air base (France). Six military teams, working in small four-person group, were trained together and required to take a large number of decisions in uncertain situations. Their activities were reproducible using a flight simulator of a BC-135 Boeing during in-flight refuelling.

Figure 3: The military air transport system cockpit.

4.1. Experimental protocol

The experimental scenario was inspired by a real incident. Initially, smoke accompanied by a burning smell is detected in the cabin. Then, a series of apparently unlinked faults occur (*e.g.*, frost on the windows, loss of fuel indications, an overheating transformer, smoke). The aircraft is over the ocean and cannot land. The problem is an electrical failure and is located in a specific area of a generator. Its fuse, which is poorly visible, has blown. In fact, all the failed components have the same origin, but expert opinion is divided between two possible causes. Thus, the team has to face an ambiguous or uncertain situation. Facing these successive faults, the team has to make sense of the situation in order to apply the correct procedures. They are not expected to know the recovery rules, but they have all the manuals with which to identify them.

Despite teams differences, comparison between teams could be achieved in terms of the criteria selected for this study, aggregated into so-called local and global resilience indicators. Several criteria were defined in order to evaluate the general Human-Machine System development: criteria related to system safety, human workload, and the team mission. These criteria are the main factors concerning system performance in the event of a major mishap and are evaluated by experts. The goal of the experiment was to observe the reaction of teams faced with unprecedented situations and to estimate the resilience indicators. The specific hypothesis to be tested during the experiment mainly concerned the way teams reacted (procedures, etc.) but is not described in this paper, since here we are focusing on the estimation of resilience indicators. A more detailed discussion and description has been realised by Ouedraogo et al. (2013).

$$U_i = \begin{pmatrix} \begin{pmatrix} u_{m\,i} \\ \sum_{i=1}^{i} u_{m\,i} \end{pmatrix} \\ \begin{pmatrix} u_{s\,i} \\ \sum_{i=1}^{i} u_{s\,i} \end{pmatrix} \\ \begin{pmatrix} u_{w\,i} \\ \sum_{i=1}^{i} u_{w\,i} \end{pmatrix} \end{pmatrix} \tag{30}$$

For iteration i, the system state in terms of local and global resilience indicators must be estimated. The resulting matrix, denoted U_I, is composed of vectors U_i depicted in Equation (30), containing the same number of rows as there are k criteria of the resilience estimation:

- The team mission criterion $u_{m\,i}$ is the percentage success in achievement of this mission.
- The safety criterion $u_{s\,i}$ relates to the recovery efficiency to faults.
- The human workload criterion $u_{w\,i}$ is linked to the number of interactions between staff members (*i.e.*, frequency of communication, actions) and between the staff and the technical system (*e.g.*, standard procedures, applied actions).

All the u_{ki} were given a score by experts between 0 and 1; with 1 indicating maximum resilience to the situation (success of mission or aircraft safe or 100% workload available) and 0 the minimum for the worst situations with respect to the resilience criteria (mission failed or aircraft crashed or overloaded operators). Then $\sum_{i=0}^{i} u_{ki}$ is assessed between 0 and i.

4.2. Reinforcement impact assessment

At iteration i, $|\mathcal{E}_{ki}|$, the absolute error for criteria k, is calculated using Equation(31).

$$|\mathcal{E}_{ki}| = \frac{1}{2} * \left(|u^*_{k\,i} - u_{k\,i}| + \frac{1}{i} * \sum_{i=1}^{i} |u^*_{k\,i} - u_{k\,i}|\right) \tag{31}$$

Then $|\mathcal{E}_i|$ is calculated to provide a single value as the error of estimation by merging all criteria with the same relative weighting. The result is in the range $[0; 1]$ in Equation (32).

$$|\mathcal{E}_i| = \frac{1}{k} * \sum_{k=1}^{k} |\mathcal{E}_{ki}| \tag{32}$$

The Mean Absolute Error (MAE) of estimation is summed for i iterations using Equation (33).

$$MAE = \frac{1}{i} * \sum_{i=1}^{i} |\mathcal{E}_i| \tag{33}$$

Table 2: Difference in MAE on the resilience indicator estimation.

	local resilience		global resilience	
Teams	ΔER	ΔKHR	ΔER	ΔKHR
Team 1	-0,29	-0,53	-0,15	-0,15
Team 2	0,05	-0,11	0,19	-0,06
Team 3	0,06	-0,20	-0,50	-0,27
Team 4	-0,13	-0,34	0,06	-0,28
Team 5	-0,28	-0,24	-0,05	-0,12
Team 6	-0,43	-0,64	-0,33	-0,49
Average	-0,17	-0,34	-0,13	-0,23

Next, the difference in MAE between two functions is calculated at the end of the experiment to evaluate the reinforcement impact on the estimation process with a result in the range $[-1; 1]$. Table 2 shows the differences in Mean Absolute Error (MAE) results obtained between no reinforcement and the Estimation reinforcement functions, denoted ΔER, and between the Knowledge reinforcement and the Hybrid reinforcement functions, denoted

ΔKHR. For the estimation of one particular team, the data from the other teams were integrated into the knowledge, because the comparisons between no reinforcement and the Knowledge reinforcement, and between no reinforcement and the Hybrid reinforcement functions could not be achieved. If the MAE difference is positive, then the first function is better than the second, *i.e.* no reinforcement gives better results than the Estimation reinforcement function for ΔER and Knowledge reinforcement is better than Hybrid reinforcement for ΔKHR. Of course, if the MAE difference is negative, then the second tested function is better than the first. ΔER was negative in each case except for $team2$. This demonstrates the effectiveness of Estimation reinforcement versus no reinforcement. For ΔKHR, the negative values, lower than ΔER in 10/12 cases, demonstrate the value of Hybrid reinforcement. Finally, the lowest average values confirm the performance of Hybrid reinforcement for both local and global resilience indicators.

Table 3: MSE on the resilience indicator estimation.

	local resilience			global resilience		
Teams	KR	ER	HR	KR	ER	HR
Team 1	1,31	0,41	0,32	0,63	0,33	0,49
Team 2	0,84	0,46	0,67	0,66	1,14	0,55
Team 3	0,70	0,88	0,40	0,76	0,43	0,43
Team 4	0,71	0,53	0,21	0,55	0,28	0,23
Team 5	0,82	0,71	0,62	0,09	0,15	0,07
Team 6	0,83	0,51	0,11	0,75	0,44	0,20
Average	0,87	0,58	0,39	0,57	0,46	0,33

The results are reinforced by Table 2, which presents the Mean Squared Error (MSE) results obtained for the Estimation Reinforcement (ER), the Knowledge Reinforcement (KR) and the Hybrid Reinforcement (HR). The same process as for MAE has been used for this calculation, *i.e.* $\mathcal{E}_{ki}$ and $\mathcal{E}_i$ were initially calculated based on errors instead of absolute errors in the range $[-1;1]$, and for the final MSE with a result between $[0;4]$ given in Equation (34).

$$MSE = \frac{1}{i-1} * \sum_{i=2}^{i} (\mathcal{E}_i - \mathcal{E}_{i-1})^2 \tag{34}$$

The function with the lowest MSE is assumed to be the most efficient, because the difference between the estimated and measured vector was minimized. A value of 0 indicates a perfect estimation of resilience indicators with no error in the three selected criteria and 4 indicates the maximum error. In Table 3, the Estimation Reinforcement function provides a 9/12 improvement over Knowledge Reinforcement, mainly because there is a lack of expert data for network initialization and training. If each team is considered independently, the performance of each function can be assessed. It can be seen that the performance of the resilience indicators are not necessarily linked; for instance, *team* 1 has a very good local HR value (less than the average MSE value) and very poor global HR value (greater than the average MSE) because the team members apply the correct manoeuvres in an incorrect sequence. the *team* 2 results are by far the worst in the experiment, because the team members did not notice smoke in the early stage of the experiment and apply the correct manoeuvres with delay. There is nothing in particular to note regarding the *team* 3 results, except that ER in local resilience is slightly better than KR but worse than HR. Finally, *teams* 1 & 2 & 3 have HR results that are above the average whereas *teams* 4 & 5 & 6 are below the average. This can be explained by the fact these teams have been asked, at the beginning of the experiment, to follow the same procedures as they would usually perform, whereas last three teams have been asked to practice new procedures. The overall performance of the teams as given by average values shows that the Hybrid Reinforcement is the most efficient function, as it combines both ER and KR for both local and global resilience indicators.

5. Conclusion

For a Human-Machine System facing a critical situation due to unprecedented events, *i.e.* events with a very rare frequency of occurrence and which may have catastrophic consequences, the concept of resilience has been defined as the positive ability to recover or to adapt to this critical situation. A brief discussion on the need to estimate a Human-Machine System state with resilience indicators has been presented. A solution employing estimation through a learning system has been selected. Therefore, two reinforced learning functions, using knowledge or estimation reinforcement, and a hybrid approach including both of these have been proposed and implemented to illustrate the feasibility of such estimations. The proposed iterative reinforced learning structure has the ability to estimate resilience indicators and to learn from experience. Because the knowledge reinforcement function requires substantial sets of representative data in order to train the network, it is less efficient than the estimation reinforcement function. The main advantage of the hybrid reinforcement function is that it combines both estimation

reinforcement and knowledge reinforcement, so that it allows the most efficient estimation results. However, it is not possible at this stage to provide consistent statistics, since only six teams were included in the experiment of this project. The system designed for the estimation of resilience indicator is a first step with the potential goal of being combined with a model of Human-Machine System, to simulate more scenario and configurations and to consolidate the results.
In the future, we aim to refine the reinforcement function in order to reduce estimation errors. For example, the relative weighting of criteria during reinforcement impact assessment could be analysed to improve results. We would also like to explore the use of more advanced machine learning methods to implement promising practical estimation structure. An example involving scenarios applied to urban guided transport systems will be addressed to perform a more in-depth analysis and to consolidate the early results.

6. Acknowledgment

The present research work has been supported by:
The Nord-Pas-de-Calais Region,
The European Community,
The Regional Delegation for Research and Technology,
The Ministry of Higher Education and Research,
The Scientific Research Group on Supervisory, Safety and Security of Complex Systems,
The International Research Network HAMASYTI (Human-Machine Systems in Transportation and Industry),

And the National Center for Scientific Research.

The authors gratefully acknowledge the support of these institutions.

7. References

Cacciabue, P. C., Enjalbert, S., Söderberg, H., Tapani, A., 2013. Unified driver model simulation and its application to the automotive, rail and maritime domains. Transportation research part F: traffic psychology and behaviour 21, 315–327.

Calinon, S., Guenter, F., Billard, A., 2007. On learning, representing, and generalizing a task in a humanoid robot. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 37 (2), 286–298.

Chella, A., Dindo, H., Infantino, I., 2006. A cognitive framework for imitation learning. Robotics and Autonomous Systems 54 (5), 403–408.

Delgado, M., Cuéllar, M. P., Pegalajar, M. C., 2008. Multiobjective hybrid optimization and training of recurrent neural networks. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 38 (2), 381–403.

Doisy, G., Meyer, J., Edan, Y., 2014. The impact of human–robot interface design on the use of a learning robot system.

Enjalbert, S., Ouedraogo, K. A., Vanderhaegen, F., et al., 2013. Validation of a unified model of driver behaviour for train domain. In: Analysis, Design, and Evaluation of Human-Machine Systems. Vol. 12. pp. 505–512.

Enjalbert, S., Vanderhaegen, F., Pichon, M., Ouedraogo, K. A., Millot, P., 2011. Assessment of transportation system resilience. In: Human Modelling in Assisted Transportation. Springer, pp. 335–341.

Geng, Z., Dong, J., Chen, J., Han, Y., 2017. A new self-organizing extreme learning machine soft sensor model and its applications in complicated chemical processes. Engineering Applications of Artificial Intelligence 62, 38–50.

Jang, T.-J., Choi, C.-H., Ahn, H.-S., 1995. Iterative learning control in feedback systems. Automatica 31 (2), 243–248.

Lee, J. H., Lee, K. S., 2007. Iterative learning control applied to batch processes: An overview. Control Engineering Practice 15 (10), 1306–1318.

Lee, T., Tan, K. K., Lim, S., Dou, H., 2000. Iterative learning control of permanent magnet linear motor with relay automatic tuning. Mechatronics 10 (1), 169–190.

Lin, C.-J., Wu, C., Chaovalitwongse, W., et al., 2015. Integrating human behavior modeling and data mining techniques to predict human errors in numerical typing. Human-Machine Systems, IEEE Transactions on 45 (1), 39–50.

Liu, D., Wang, D., Yang, X., 2013. An iterative adaptive dynamic programming algorithm for optimal control of unknown discrete-time nonlinear systems with constrained inputs. Information Sciences 220, 331–34lopez2002knowledge2.

Norrlöf, M., Gunnarsson, S., 2005. A note on causal and cite iterative learning control algorithms. Automatica 41 (2), 345–350.

Ojha, V. K., Abraham, A., Snášel, V., 2017. Metaheuristic design of feedforward neural networks: A review of two decades of research. Engineering Applications of Artificial Intelligence 60, 97–116.

Orwin, K. H., Wardle, D. A., 2004. New indices for quantifying the resistance and resilience of soil biota to exogenous disturbances. Soil Biology and Biochemistry 36 (11), 1907–1912.

Ouedraogo, K. A., Enjalbert, S., Vanderhaegen, F., 2013. How to learn from the resilience of human–machine systems? Engineering Applications of Artificial Intelligence 26 (1), 24–34.

Pérez-España, H., Arreguın-Sánchez, F., 2001. An inverse relationship between stability and maturity in models of aquatic ecosystems. Ecological Modelling 145 (2), 189–196.

Polet, P., Vanderhaegen, F., Millot, P., 2009. Human behaviour analysis of barrier deviations using a benefit-cost-deficit model. Advances in Human-Computer Interaction 2009, 4.

Polet, P., Vanderhaegen, F., Zieba, S., 2012. Iterative learning control based tools to learn from human error. Engineering Applications of Artificial Intelligence 25 (7), 1515–1522.

Radac, M.-B., Precup, R.-E., 2016. Three-level hierarchical model-free learning approach to trajectory tracking control. Engineering Applications of Artificial Intelligence 55, 103–118.

Rose, L., Perlaza, S. M., Le Martret, C. J., Debbah, M., 2014. Self-organization in decentralized networks: A trial and error learning approach. Wireless Communications, IEEE Transactions on 13 (1), 268–279.

Vanderhaegen, F., 1997. Multilevel organization design: the case of the air traffic control. Control Engineering Practice 5 (3), 391–399.

Vanderhaegen, F., 2010. Autonomy control of human-machine systems. IFAC Proceedings Volumes 43 (13), 398–403.

Vanderhaegen, F., Chalmé, S., Anceaux, F., Millot, P., 2006. Principles of cooperation and competition: application to car driver behavior analysis. Cognition, Technology & Work 8 (3), 183–192.

Vanderhaegen, F., Polet, P., Zieba, S., 2009. A reinforced iterative formalism to learn from human errors and uncertainty. Engineering Applications of Artificial Intelligence 22 (4), 654–659.

Vanderhaegen, F., Zieba, S., 2014. Reinforced learning systems based on merged and cumulative knowledge to predict human actions. Information Sciences 276, 146–159.

Vanderhaegen, F., Zieba, S., Enjalbert, S., Polet, P., 2011. A benefit/cost/deficit (bcd) model for learning from human errors. Reliability Engineering & System Safety 96 (7), 757–766.

Wiering, M. A., van Hasselt, H., 2008. Ensemble algorithms in reinforcement learning. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 38 (4), 930–936.

Xu, J.-X., Lee, T. H., Zhang, H.-W., 2004. Analysis and comparison of iterative learning control schemes. Engineering Applications of Artificial Intelligence 17 (6), 675–686.

Xu, J.-X., Yan, R., 2004. Iterative learning control design without a priori knowledge of the control direction. Automatica 40 (10), 1803–1809.

Yan, M.-T., Shiu, Y.-J., 2008. Theory and application of a combined feedback–feedforward control and disturbance observer in linear motor drive wire-edm machines. International Journal of Machine Tools and Manufacture 48 (3), 388–401.