A neural network encoder-decoder for time series prediction: Application on 137Cs particulate concentrations in nuclearized rivers

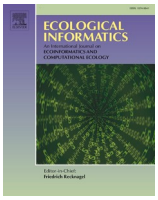
Kathleen Pele, Valérie Nicoulaud-Gouin, Hugo Lepage

Ecological Informatics 80 (2024) 102463

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



A neural network encoder-decoder for time series prediction: Application on ^{137}Cs particulate concentrations in nuclearized rivers

Kathleen Pelé *, Valérie Nicoulaud-Gouin, Hugo Lepage

Institut de Radioprotection et de Sûreté Nucléaire (IRSN), PSE-ENV/SRTE/LRTA, F-13115, Saint-Paul-lez-Durance, France

ARTICLE INFO

Keywords:
Deep learning
Suspended particulate matter
Radioactivity
Micropollutant
Data-driven model

ABSTRACT

Monitoring the impact of human activities on the environment is a major challenge as many pollutants can be found in the different ecosystems. This is the case of the caesium-137 that has been present in the environment for many decades as a result of atmospheric tests, accidents such as Chernobyl and release from nuclear industries. With the recent advance in data-driven models, this study evaluate the relevance of a deep learning tool for reconstructing caesium-137 chronics particulate concentration in rivers. An encoder-decoder neural network, "Hierarchical Attention-Based Recurrent Highway Networks"(HRHN), is proposed notably for its ability to extract the most relevant temporal and spatial information from the databases. Three monitoring stations were studied, one on the Rhône River and two on the Loire River, all of them downstream nuclear industries in these catchments affected by the global fallout and the accident of Chernobyl. The objective is to predict the future concentration from a set of variables providing past information on water discharge, washout flux and industrial radioactive releases. Once optimised, the model generates first results in agreement with the real concentration curves by correctly following the main trends, with a NSE of 0.89, 0.53 and 0.35 respectively for the Rhone station and the two stations on the Loire. The main reason of inaccuracies is due to the quantity of data available. The originality of this model is its capacity to make predictions on different catchment areas. In fact the training was conducted on the Rhône station as the range of the concentration was higher (from 265.4 to 2700.0 Bq/kg) and the testing on the two Loire station. Another encoder-decoder model DA-RNN (Dual-Stage Attention-Based Recurrent Neural Network) was also evaluate in order to compare the performance of an alternative architecture, without convolution layer. The conclusion is that HRHN remains more powerful in the predictions on the 3 systems. With these first interesting results for HRHN, further investigations should be taken into account for other pollutants than caesium-137 to better understand the robustness of the model.

1. Introduction

The presence of artificial radioactive materials in the environment has been proved since decades, particularly following atmospheric nuclear testing, and accidents such as Fukushima and Chernobyl (Tracy et al., 2013). Understanding the fate of these contaminants and their resiliencies is therefore essential as most of them have a long half-time and remains on ecosystems especially those coming from the nuclear industries (Hirose and Povinec, 2022; Kashparov et al., 2019). This is the case for caesium-137 (^{137}Cs), which will persist in the soil affected by these fallouts in catchment areas for 30 years and contaminate aquatic environments through the erosion of soil (Lepage et al., 2016). Moreover, ^{137}Cs is also emitted into rivers during authorized discharges from nuclear facilities (Eyrolle et al., 2020a). Therefore, its behaviour in hydrosystems has been studied for many years (Konoplev et al., 2020; Kryshev, 1995; Takahashi et al., 2017; Yoshimura et al., 2015) and a large number of models have been developed to understand its transport in watercourse (Ikenoue et al., 2023; Iwasaki et al., 2015; Konoplev et al., 2020; Tomczak et al., 2021).

The recent development in artificial intelligence methods and their implementation in several libraries (e.g. scikitlearn library on Python) (Oludare Isaac et al., 2018) brings new tools able to estimate the concentration of various pollutants or hydrological quantities by using existing databases (Yaseen, 2021). Indeed, a number of these methods are effective in identifying non-linear relationships or structures in the data allowing to realize predictions on the quantity studied with a good accuracy and a high speed of execution. In recent years, there has been a significant number of articles with a review of machine learning and

* Corresponding author.
E-mail addresses: kathleen.pele@irsn.fr (K. Pelé), valerie.nicoulaudgouin@irsn.fr (V. Nicoulaud-Gouin), hugo.lepage@irsn.fr (H. Lepage).

https://doi.org/10.1016/j.ecoinf.2024.102463
Received 30 August 2023; Received in revised form 3 December 2023; Accepted 6 January 2024
Available online 9 January 2024

deep learning methods applied to river monitoring and water quality (Rajaee et al., 2020). For examples on neural network methods, Adaptive Neuro-Fuzzy Inference Systems (ANFIS) have been applied to various domains of water quality forecasting (Tiwari et al., 2018). A MultiLayer Perceptron (MLP) gives satisfactory responses in the modelling of pH with hydrometeorological data such as discharge and solar radiation (Moatar et al., 1999) and a SVM (Support Vector Machine) were used to predict the nitrate concentration in river water (Stamenković et al., 2020). A decision tree model was applied for the forecast of sediment yield generated within a watershed (Goyal, 2014). However, there are very few papers applying these methods to the prediction of radionuclide concentrations in the environment (Dragović, 2022). A few attempts have been made, such as (Shuryak, 2022) who uses RF (Random Forest) to predict the concentration of ^{137}Cs in terrestrial plants, or (Kulahci et al., 2006) which constructs an MLP for the prediction of two outputs on alpha and beta radioactivity. To our knowledge, there is no study on the use of data models to predict ^{137}Cs concentrations in a nuclearized river. Notably, there are no deeper networks applied to this topic or the use of the temporality of the data because these networks are sometimes developed in communities outside environmental sciences.

The development of neural networks applied to time series has grown rapidly in recent years especially encoder-decoder network (EDN) (Cho et al., 2014a, 2014b). The basic principle of EDNs is to compress the information of exogenous time series into a latent representation (to highlight important characteristics) via the encoder, and then decompress it to predict future values via the decoder. The nature of the layers composing the encoder and decoder can be of different types (convolutional, recurrent, …). For example, in the paper (Fawaz et al., 2019), the authors present a convolutional encoder to extract features from the time series and then a convolutional decoder to reconstruct the series from these features for time series classification. In (Yang et al., 2019), the authors propose an EDN architecture based on a convolutional encoder and an recurrent decoder for traffic flow prediction. EDNs that combine convolutional and recurrent layers allow efficient modelling of time sequences for prediction. This combination allows robust features to be extracted from the time series and longterm relationships in the sequence to be modelled, which is particularly important for time series prediction tasks. A other work on time series prediction is the DA-RNN (Dual-Stage Attention-Based Recurrent Neural Network) model (Qin et al., 2017), the encoder exploits an input attention mechanism to adaptively extract relevant input features at each time step by referring to the previous encoder's hidden states. And the decoder uses a classical attention mechanism to select relevant encoder's hidden states across all time steps. These two attention mechanism's are well integrated within a recurrent neural network.These attention mechanisms allow the most important parts of the time sequence to be highlighted. This is particularly useful for complex time series, where models may need to understand the relationship between distant elements in the sequence. The article (Tao et al., 2016) provides an architecture which combines a set of very interesting aspects (of the previously mentioned architectures): use of convolution, recurrent networks and an attention mechanism. The proposed architecture is the encoder-decoder neural network Hierarchical attention-based Recurrent Highway Networks (HRHN), a model that has not yet been seen in the environmental literature (Chen and Li, 2020; Shoham and Permuter, 2018; Zhang et al., 2017) and has several advantages over other neural networks, including:

- Attention hierarchy: HRHN uses an attention hierarchy that identifies the most important parts of the inputs at each hierarchical level, which can improve prediction performance.
- Recurrence and long-term connectivity: HRHN uses recurrent layers that allow for long-term memory of previous inputs
- Convolution layers: the use of convolution layers allows the detection of local patterns in the time series and the extraction of relevant features at different time scales.
- Use of highway gates: HRHN uses highway gates that allow the amount of information that is passed between layers to be controlled, which can facilitate the training of complex models.

The ambition of this study is to propose a model based on the HRHN that can understand the dynamic of the ^{137}Cs concentration in the river with different sources: accidental events (like Chernobyl) or standard events (chronic releases from nuclear facilities). Thereafter, the aim is to predict the future concentration of ^{137}Cs in SPM (Suspended Particulate of Matter) for different rivers from a set of variables providing past information on water discharge, washout flux and release data. The paper is organized as follows. The section 2 of this paper presents the different data sets studied for the rivers and the temporal modelling strategy chosen. The section 3 presents the HRHN neural network. Section 4 and 5 presents respectively the optimization of the model and the sensitivity analysis of HRHN. The section 6 presents an alternative to HRHN, the DA-RNN model. Section 7 presents respectively the most relevant results and a discussion on the results. Finally, the article ends with the section conclusion in section 8.

2. Material and methods

2.1. Presentation of the studied watersheds

Datasets from two rivers are used in this study: the Rhône and the Loire Rivers. They are the most nuclearised rivers in France with several nuclear industries (Eyrolle et al., 2020b; Goutal et al., 2008) among them 5 nuclear power plants (NPP) and a nuclear waste reprocessing site on the Rhône River and 5 nuclear power plants (NPP) on the Loire River. These two rivers are also the receptacle of artificial radionuclides drained from soils mainly marked by atmospheric fallout from military nuclear tests (between 1945 and 1980) and the Chernobyl accident (1986)(Meusburger et al., 2020; Roussel-Debel et al., 2007).

The Rhône watershed (97,800 km^2) is characterised by a strong climatic and geological heterogeneity that leads to a strong variation of annual SPM fluxes (from 1.4 Mt. to 18.0 Mt./year (Delile et al., 2020; Poulier et al., 2019)), while the Loire basin (117,500 km^2) has an annual variation in these flux ranging from 0.3 to 1.2 Mt./year (Moatar and Dupont, 2016).

Three monitoring stations are considered, one on the RhÃ'ne River and two on the Loire River (Fig. 1).

- Rhône-Vallabregues: the station of Vallabrègues downstream the Marcoule reprocessing center from 1983 to 04-01 to 2007-01-01
- Loire-Ouzouer: the station of Ouzouer in the Loire River (downstream the Dampierre and Belleville NPP) from 1987 to 06-01 to 2006-12-01
- Loire-Muides: the station of Muides in the Loire River (downstream the Saint-Laurent-Des-Eaux, Dampierre and Belleville NPP) from 1987 to 06-01 to 2006-12-01

2.2. Variables of the three systems

In order to predict the ^{137}Cs concentration variable (endogenous variable) in suspended particulate matters, the hydrological characteristics (water discharge) and the ^{137}Cs sources (quantity of nuclear industry release and washout flux) were considered as exogenous variables. All the chronicles studied have a monthly time step. The endogenous variable is the variable to be predicted. The exogenous variables are to explain the behaviour of the endogenous variable.

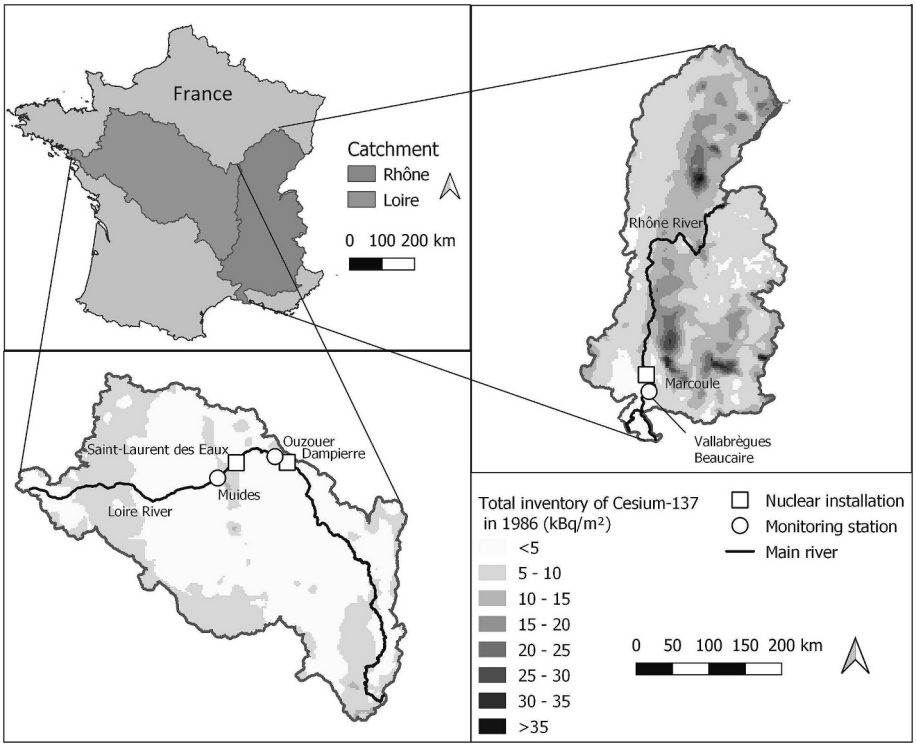


Fig. 1. Location of the studied monitoring stations in the Rhône and Loire Catchments. Total inventory of ^{137}Cs was estimated by (Roussel-Debel et al., 2007).

2.2.1. Suspended particulate matter data

Measurements of ^{137}Cs are taken from the monitoring of releases from the facilities set up in the 1980s. Sediment traps were used to collect SPM every month then SPM samples were slowly (approx. 2 weeks) evaporated (80 °C) to dryness, ashed and put into tightly closed plastic boxes (17*mL* or 60*mL*) for gamma-ray spectrometry measurements (20–60 g) using low-background and high-resolution (High Purity Germanium detectors). Results are expressed in Bq/kg and each sample was measured for 3 days to achieve detection limits around 0.5 Bq/kg.

The Fig. 2 shows the different trends in concentrations according to the sites studied. Over the period studied (1983–2007), the average concentration of ^{137}Cs in the Rhône-Vallabregues station was 265.4 Bq/kg, with a maximum of 2700.0 Bq/kg in December 1985 and a minimum of 4.8 Bq/kg in April 2001. The Loire-Ouzouer system over the period studied (1987–2007) had an average ^{137}Cs concentration of 15.1 Bq/kg, with a maximum of 130 Bq/kg in September 1987 and a minimum of 3.2 Bq/kg in August 2006. The Loire-Muides station over the period studied (1987–2007) had an average ^{137}Cs concentration of 21.3 Bq/kg,

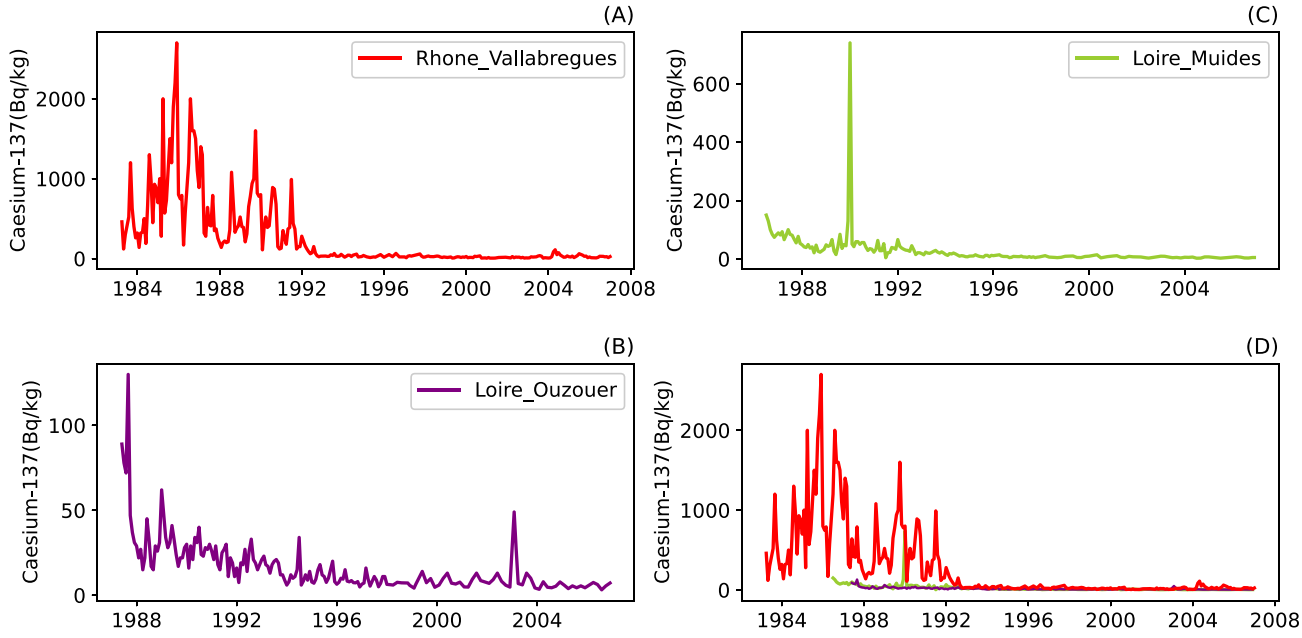


Fig. 2. Concentration of ^{137}Cs in SPM collected in the three studied stations, (A) Rhône-Vallabregues, (B) Loire-Ouzouer, (C) Loire-Muides and (D) the three compiled.

with a maximum of 740 Bq/kg in January 1990 and a minimum of 2.3 Bq/kg in July 2004.

2.2.2. Water discharge data

The discharge values for the three monitoring stations (Fig. 3) were obtained from the "HydroPortail" national database (https://hydro.eaufrance.fr/rechercher/entites-hydrometriques) and are expressed in m^3/s. They are collected at the station closest to the sampling site: for Rhone-Vallabregues, the hydrometric site is Beaucaire (code site: V720 0005), for Loire-Muides, the hydrometric site is Blois (code site: K447 0010) and for Loire-Ouzouer, the hydrometric site is Giens (code site: K418 0010). The data is retrieved in daily format and then transformed into monthly data to respect the scale of other variables by selecting the maximum and minimum dischage water during the month and the monthly average. The Fig. 3 shows the monthly minimum discharges water for each site (graphs (A),(B) and (C)) and a comparison of the minimum discharges water for the 3 sites (graph (d)). To improve the visibility of the graph, only the minimum discharge is shown. For the Rhône-Vallabregues station over the period studied (1983–2007), the average water discharge is 1694.8 m^3/s, with the highest discharge at 10900 m^3/s in December 2003 and the lowest discharge at 322 m^3/s in January 1990. For the Loire-Ouzouer station over the period studied (1987–2007), the average water discharge is 310.7 m^3/s, with a maximum discharge of 3130 m^3/s in December 2003 and a minimum discharge of 30.8 m^3/s in August 1991. For the Loire-Muides station over the period studied (1987–2007), the average dischagre is 329.4 m^3/s, with a maximum discharge of 2940 m^3/s in December 2003 and a minimum discharge of 28 m^3/s in July 2006.

2.2.3. Release data

The NPP are allowed to release radioactive effluent directly into river (Eyrolle et al., 2020b). Such release must respect concentration thresholds and be carried out under normal hydrological conditions (baseflow), excluding low-level water and flood. For ^{137}Cs, the main source of liquid effluent in the Rhône River is the reprocessing center of Marcoule which represent most of the annual emission. The installation of the liquid effluent treatment (STEL) plant at the Marcoule site earlier in the 90s has significantly reduced the quantity releases (Fig. 4-(A)). The STEL is responsible for reception, treatment of liquid radioactive effluents from the site and discharge of decontaminated liquid effluent into the Rhône River, after filtration and control. In the Loire River the releases come only from the NPPs and only the following NPP were studied Dampierre, Belleville Saint-Laurent-des-Eaux. To be more precise, for Loire-Ouzouer this is the chronicle of cumulative releases from Dampierre and Belleville and for Loire-Muides the chronicle of cumulative release from Saint-Laurent-des-eaux, Dampierre and Belleville. The Fig. 4 shows the different trends in releases according to the sites studied. For the Rhône-Vallabregues over the period studied (1983–2008), the average release was 59,997.6 MBq, with a maximum release of 360,000 MBq in June 1985 and a minimum release of 476 MBq in August 2002. For the Loire-Ouzouer station over the period studied (1987–2007), the average release was 91.3 MBq, with a maximum release of 1500 MBq in August 1987 and a minimum release of 2.71 MBq in August 1991. For the Loire-Muides station over the period studied (1987–2007), the average release was 112.9 MBq, with a maximum release of 1564 MBq in August 1987 and a minimum release of 3 MBq in February 2006.

2.2.4. Washout flux data

The purpose of this variable is to represent the pollution from phenomena outside the river. Washout refers to the washing of the atmosphere and the soil during rainfall (Borzilov et al., 1988; Khanbilvardi et al., 1999). This phenomenon pollutes run-off water and contaminates rivers. Raw washout data is not available. This is why we propose to estimate he data on the ^{137}Cs washout flux by the work done by (Vrel, 2012). This washout flux is expressed in MBq/s and is obtained by the convolution product of the atmospheric deposition flux and a transfer function (characteristic response of the river after a point contamination) (Delmas et al., 2017; Vrel, 2012). It should be noted that this calculation was made using atmospheric data from the Seine River, a large french catchment, as data specific to the Rhône and Loire Rivers were not available. The contribution of bombs is negligible because they are outside our study period (or at the very end for Marcoule, where releases mask them). It is assumed that the impact of Chernobyl on all French rivers may result in a peak in 1986 of washout fluxes and then decrease (Roussel-Debel et al., 2007). Only the amplitude of this peak could vary according to the river. Inventory of ^{137}Cs in soil at the date of 1986 were used to correct the flux data (Roussel-Debel et al. (2007), Fig. 1). The estimated inventories in Bq were respectively for the Seine, the Rhône and the Loire Rivers 3.9473×10^{14}, 9.3965×10^{14} and 5.73821×10^{14}. Therefore, the flux data estimated from the Seine were

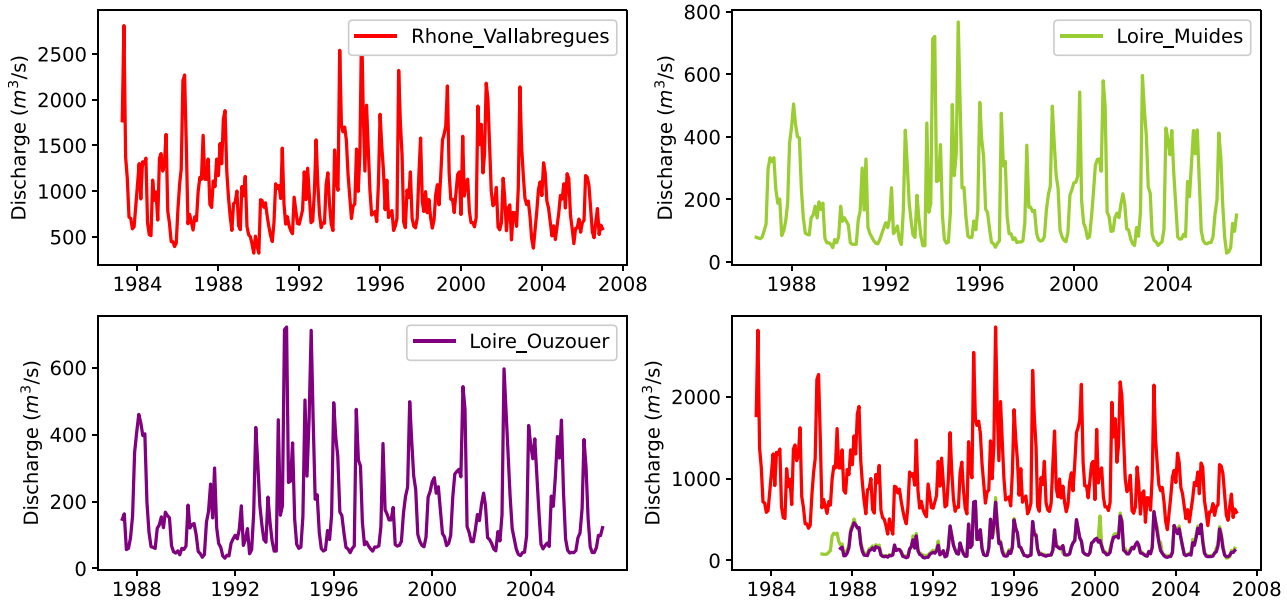


Fig. 3. Water discharge (min) measured (A) Rhône, (B) Loire-Ouzouer, (C) Loire-Muides and (D) comparison of minimum discharge water for the 3 sites.

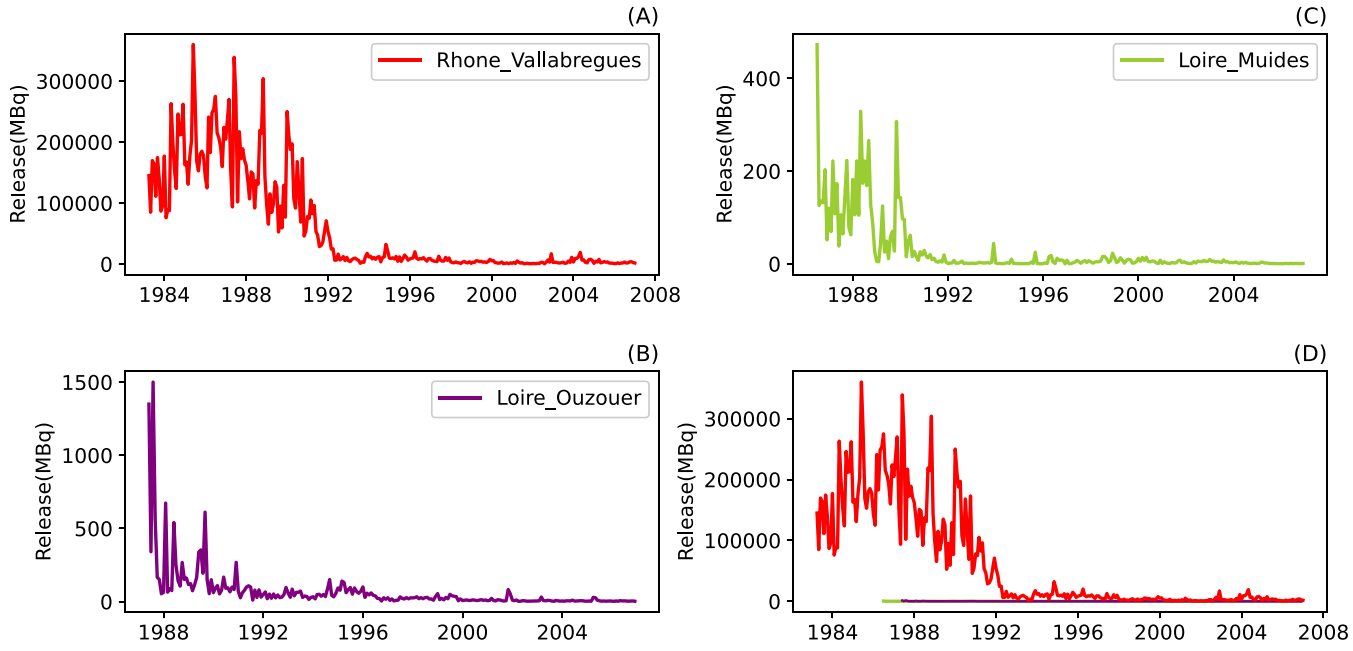


Fig. 4. Releases measured for (A) Rhône-Vallabregues (Marcoule reprocessing center), (B) Loire-Ouzouer (the Dampierre NPP and Belleville NPP), (C) Loire-Muides (the Dampierre NPP, Belleville NPP and Saint-Laurent-Des-Eaux NPP) and (D) the three compiled.

multiplied by 138% in the case of the Rhône River and by 45% in the case of the Loire River. In fact, while the fallout from the nuclear test was almost homogeneous in France, the Chernobyl fallout was more important in the Eastern France (including the Rhône Catchment) than the Western France (Fig. 1). The Fig. 5 shows the different trends in concentrations according to the sites studied. The peak of washout flux was respectively 125 MBq/s and 5.5 MBq/s for the Rhône-Vallabregues and the Loire-Ouzouer stations while the last values concidered were lower than 0.05 MBq/s.

2.2.5. *Selected variable and transformation*

Three explanatory variables are retained: washout flux, releases and minimum discharge water. More details on the choice of these variables are explained in section 5. The data have been transformed using the PowerTransform library (Yeo and Johnson, 2000) because of the high level of asymmetry in the distribution of the data especially for the releases and washout flux variables. Power transforms are a technique for transforming numerical input or output variables to have a more Gaussian probability distribution. In addition, this transformation of the data also provides systems on a more comparable scale, particularly with regard to the very high releases from the reprocessing center. This is often described as removing a skew in the distribution, although more generally described as stabilizing the variance of the distribution. Power Transformer is used with the Yeo-Johnson transform. The optimal

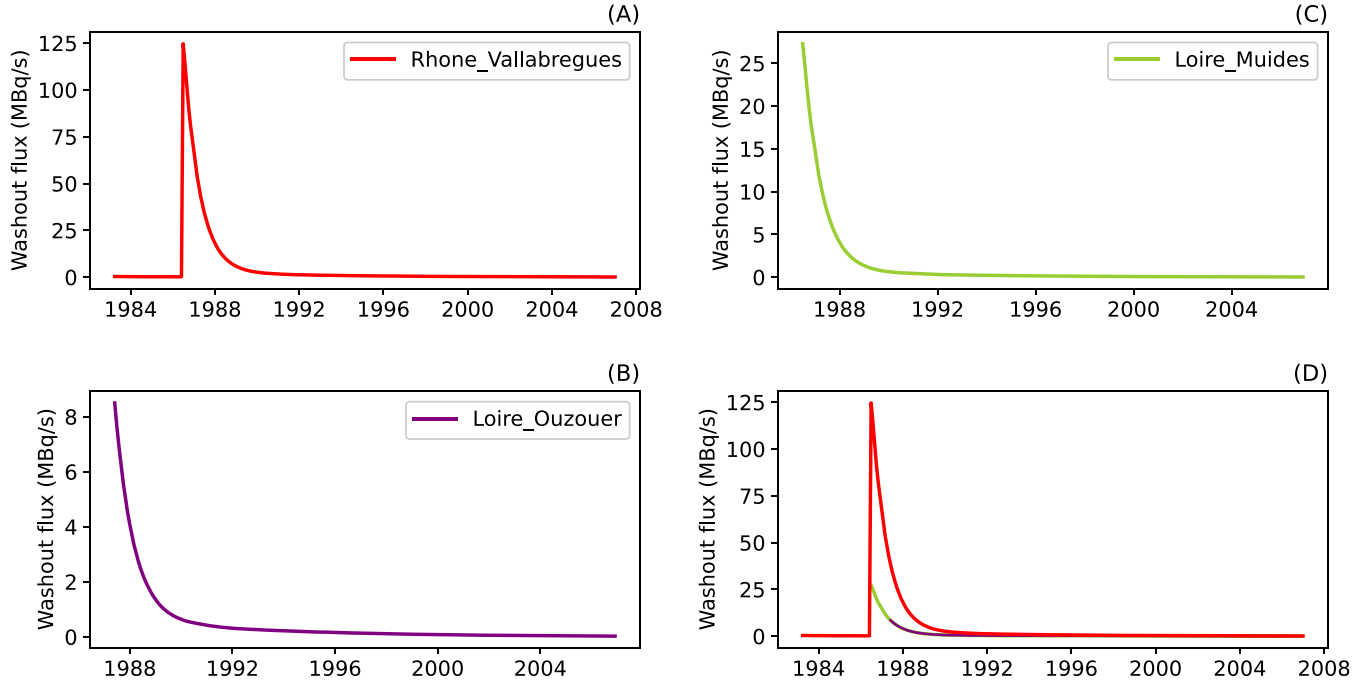


Fig. 5. Washout flux estimated for (A) Rhône, (B) Loire-Ouzouer, (C) Loire-Muides and (D) the three compiled.

parameter for stabilizing variance and minimizing skewness is estimated through maximum likelihood.

This transformation, followed by a standardisation, will be applied when training the model. The transformation and standardisation parameters are calculated from selected training base.

2.2.6. Problem formalisation and modelling choices

In order to formalise the problem, let $\left(x_t^j\right)_{t\in[1,T],j\in[1,3]}$ the exogenous series and $\left(y_t\right)_{t\in[1,T]}$ the endogenous series where t is the monthly time step between the beginning and the end of the chronicle and j the number of the exogenous variables. Time series can be modelled by F using a number of techniques, including:

- *Regression models* fit a mathematical function to the observed data to predict the values of the variable to be predicted from the values of the explanatory variables(ex. linear regression, logistic regression).
- *Smoothing models* use a weighted combination of past observations to predict future values (ex. single exponential smoothing).
- *ARIMA method* used to model stationary time series. They use the structure of the time series to determine the relationships between past and future values.
- *Neural network models* is a mathematical model composed of layers of interconnected artificial neurons that are able to learn from examples provided in the training phase. There are several types of neural networks, including feedforward neural networks, recurrent neural networks and convolutional neural networks.

In our framework, neural networks are chosen for our modelling for several reasons. The objective of time series forecasting is to generate the future series y_T based on the historical observations $y_1, y_2, \cdots, y_{T-1}$. However, the observations y_T are often related to some exogenous variables $\left(x_t^j\right)_{j\in[1,3],t\in[1,T-1]}$. For this reason, different models have been proposed for time series prediction with access to the exogenous data. The objective is therefore to determine the model F for predict the time serie of the ^{137}Cs concentration at time T from history of exogenous variables $(x_1, \ldots, x_{T-1})$ and the target variable $(y_1, \ldots, y_{T-1})$:

$$y_T = F((x_1, \ldots, x_{T-1}), (y_1, \ldots, y_{T-1}))$$

By their training capacity, neural network can learn complex patterns from the data, including non-linear relationships between variables. Neural network models can also be configured to take into account specific characteristics of the time series, such as trends, seasonality and cycles. Because of the adaptability of neural networks, they can be trained on real-time data, allowing them to adapt quickly to changes in data patterns.

3. Neural network for time serie prediction: Hierarchical attention-based Recurrent Highway Networks (HRHN)

The article (Tao et al., 2016) provides an architecture which combines a set of very interesting aspects: use of convolution, recurrent networks and an attention mechanism. The proposed architecture is the encoder-decoder neural network Hierarchical attention-based Recurrent Highway Networks (HRHN).

3.1. Architecture of HRHN

The different layers of the HRHN architecture are detailed in this section (Fig. 6):

- **Encoder with:**
  - **Convolutional network 1D (CNN-1D)** (Lecun and Bengio, 1995): to learn the spatial relationships between the different values of the exogenous series. Applying a 1D convolution involves sliding a convolution kernel over the input data by multiplying the values of the kernel with the values in the series. The aim is to detect different patterns and structures in the series (seasonal events, non-perceptible peaks in the raw series). The key strength of CNN is that it automatically learns the feature representation by convolving the neighboring inputs and summarizing their interactions. Max pooling (Aggarwal, 2018) is also performed between successive convolutional layers, which can reduce the size of feature maps so as to avoid overfitting and improve efficiency.
  - **Recurrent Highway Network (RHN)** (Zilly et al., 2016): to learn the temporal relations between the results of the previous among convolved input features. The main idea behind recurrent networks is to use an internal memory to store information about previous processing steps. In this way, the network can take into account the historical context to better understand the inputs and produce more accurate outputs. However, classical recurrent networks can encounter difficulties when training long sequences due to the "vanishing gradient" phenomenon. Indeed, the longer the sequence, the more the backpropagation of the error can become diluted and no longer effective in adjusting the network parameters. To solve this problem, other recurring structures have been proposed like RHN. The RHN network is an extension of the LSTM ((Long-Short-Term-Memory) networks. Highway connections allow data to "skip" layers that do not contribute significantly to the final prediction, which facilitates training by allowing gradients to propagate more easily through the layers. More specifically, the RHN uses gates to control the flow of data between the layers of the network. These gates are non-linear functions that take as inputs the outputs of the previous layer and the outputs of the current layer, and determine how much data should be passed to the next layer. By using these gates, the RHN can learn to "ignore" some layers that are not useful for the final prediction. The transformation gate acts as a selection and control of information from history, and a transport gate can transport information between hidden states without any activation function (Schmidhuber, 1992).
- **Attention mechanism**:

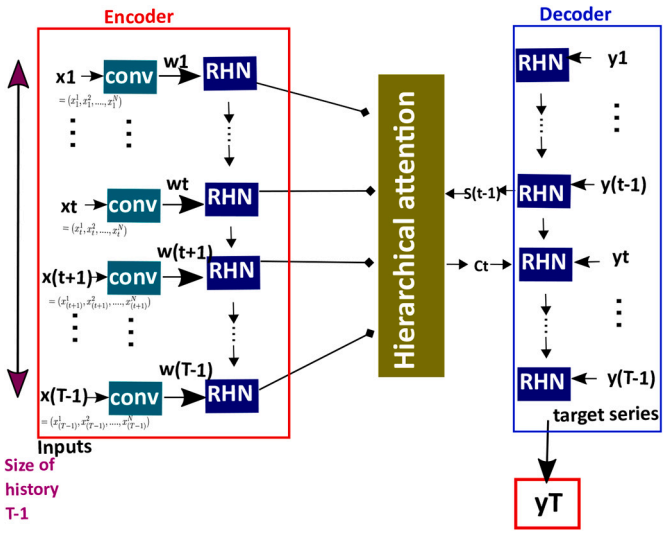


Fig. 6. A graphical illustration of HRHN. In the encoder, convolution extracts the "spatial" information of the exogenous inputs in time slice format (set of exogenous series taken at time t). Then an RHN reads the convolved features $(w_1, w_2, \ldots, w_{T-1})$ and models their temporal dependencies at different semantic levels. Using a hierarchical attention mechanism, the decoder selects the most relevant spatio-temporal features of exogenous data. The context vector c_t that feeds into the decoder RHN is obtained by concatenating all attentions. The decoder RHN capture the long-term dependencies of target series and produce the future prediction y^T (Tao et al., 2016).

– **Attention mechanism (**Bahdanau et al., 2014**):** is a neural layer allowing the model to learn to weight the information provided by the encoder according to their importance at a given time in relation to the decoding of the target series. The model learns to focus on the most relevant elements in the input sequence for the prediction task.
- **Decoder with:**
– **Recurrent Highway network (RHN) (**Zilly et al., 2016**):** of the same structure as the one present in the encoder

For more details on the theory of the elements discussed we refer the reader to the reference book (Goodfellow et al., 2016). This architecture involves a large number of hyperparameters related to the different types of layers (convolutional or recurrent) and therefore a large number of parameters. As a reminder, a parameter is internal to the neural network. It will evolve during the whole training process A hyperparameter is external to the training process, it defines the properties of the network. It remains static during the training process.

The HyperBand algorithm is proposed for optimisation and is described in the next section.

During the training of the model, the data is then standardized. In section 4, one of the three datasets will be defined as a training base, on which the standardizaion parameters will be calculated and applied to the other two datasets considered as test base. The best result of the optimisation will allow us to determine the base that will be used for training.

4. Model optimization: Determination hyperparameters and parameter training

4.1. Optimisation algorithm and experimental design

Unlike the original article (Tao et al., 2016), an optimisation is perform to determine the best set of hyperparameter values. Optimising hyperparameters is an iterative process that often involves testing several combinations of values for each hyperparameter to find the best values for the specific task. The algorithm used is the Hyperband (Li et al., 2018) method inspired by multi-armed bandit problem (available on the Keras library). The Hyperband algorithm is a hyperparameter optimisation method based on the “successive halving” strategy. Here is a brief explanation:

1. Initial sampling: Hyperband begins by randomly sampling a set of hyperparameters for model architectures.
2. Partial training: Models are partially trained (on a small fraction of the data) to quickly eliminate underperforming configurations.
3. Successive Halving: The remaining configurations are grouped into sets of different sizes, and the associated models are trained further. The best-performing configurations in each set are promoted to the next stage, while the under-performing configurations are eliminated.
4. Repeat: Steps 2 and 3 are repeated until only one configuration remains, which is then considered the best configuration found.

The key idea behind Hyperband is to explore several configurations in parallel while allocating more training resources to promising configurations. This enables a more efficient search of the hyperparameter space, particularly when computational resources are limited. In summary, Hyperband combines an initial random search with a ‘successive halving’ strategy to quickly identify promising configurations while eliminating those that show inferior performance.

The detail of the grid of possible combinations is presented. The number of convolution layers is lower than that given in the article, the network has two convolution layers. Their size and the associated max-pooling will be determined in the following interval for each:

- *CNN window size* (dim-filter-cnn) $\in [3,4,5,6,7]$. The convolution window size determines the size of the region over which convolution is applied at each time step. It is important to choose an appropriate window size to capture the relevant temporal patterns in the data.
- *the number of filters* (nbr-filters-cnn) $\in [8,16,24,32,40,48,56,64,72,80,88,96,104,112,120,128]$.The number of filters determines how many different patterns the network can learn. The higher the number of filters, the more complex the network can be, but this can also make training more difficult.
- *max pooling size*(dim-max-pooling) $\in [2,3,4]$. The pooling window size determines the region of the input that will be aggregated into a single output element. In general, a larger pooling window size reduces the spatial resolution of the output, but can also improve the robustness of the network to minor variations in the input. In contrast, a smaller pooling window size retains more detail of the input, but may also make the network more sensitive to noise or minor variations.

As in the article(Tao et al., 2016), the RHN has same structure in the encoder and the decoder:

- *hidden layers* (nbr-layers-RHN) $\in [1,2,3,4,5]$. The hidden layers allow the neural network to model non-linear relationships between inputs and outputs. Each hidden layer in a deep neural network computes a non-linear transformation of the previous layer's outputs, allowing the network to learn increasingly abstract and complex features as information is propagated through the network.
- *dimension of hidden state* (dim-RHN) $\in [8,16,24,32,40,48,56,64,72,80,88,96,104,112,120,128]$.The dimension of the hidden state determines the size of the hidden state vectors that are calculated at each time step of the model. A higher dimension of the hidden state can allow the model to capture more complex and subtle information in the data, but it can also make the model slower to train and require more training data.

The intervals chosen for these different hyperparameters are based on the following references (Goodfellow et al., 2016), (Chollet, 2018) and documentation available on Tensorflow. The algorithm Hyperband has been customised to include optimisation of the number of time steps. It's implies that for each combination of selected hyperparameters an update of the data size is performed. This hyperparameter linked to the data history allows us to determine the quantity of past information most relevant to predict the future evolution of the endogenous variable. The time step (here monthly) is selected between 3 and 20 months (lenght-sequence).

The aim of the Hyperband algorithm is to propose the most relevant hyperparameters and parameters for the model according to a training base. It therefore provides a parameterised and hyperparameterised model as output.

For these studies, it is proposed to use the MAE (mean absolute error) metric and the NSE (Nash-Sutcliffe efficiency) score for the training phase:

$$NSE = 1 - \frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{T}(y_t - \bar{y}_t)^2} \quad \text{et} \quad MAE = \frac{\sum_{t=1}^{T}(|y_t - \hat{y}_t|)}{T}$$

Two elements will be determined here:

1. The best training base between Rhône-Vallabregues, Loire-Ouzouer and Loire-Muides
2. The optimal size of the selected training base

4.1.1. Influence of the training base

The influence of the training base on the results is seen by swapping

the role of training base and test base of each system is swapped. The Hyperband algorithm is applied to each database to determine the best model (hyperparameters + parameters). This first treatment will allow us to identify the most suitable database to be the training base.

The Table 1 shows the results obtained. The use of the Rhône-Vallabregues station as training base gives the best results. The Loire-Ouzouer and Loire-Muides stations used as a training base are unable to make good predictions on the Rhône-Vallabregues station. Indeed, their informations are relatively poor to allow a relevant prediction on Rhône-Vallabregues station. Moreover, the predictions on Loire-Ouzouer and Loire-Muides stations are relatively stable whatever the training base used. It should be remembered that the aim of the optimisation algorithm is to determine the combination of hyperparameter and parameter to minimise the error on the training base and test bases. This is why when Loire-Ouzouer station is used as a training base, the predictions on its own system are not necessarily improved. Furthermore, the ^{137}Cs concentration time series of the Loire-Muides station is characterised by a very strong concentration peak in January 1990 which differs from the rest of the values taken by the series.This spontaneous and intense information is difficult to reproduce accurately in the 3 cases presented. Therefore, the prediction performance for this system is limited. The Rhône-Vallabregues station is retained as the most relevant training base. The hyperparameters and parameters resulting from the calculation with Hyperband are also retained. The selected hyperparameters are presented in the Table 2. The Adam minimisation algorithm is used with a training rate of 0.003.

4.1.2. *Influence of the size of the training base*

The size of the selected base (Rhône-Vallabregues) and its impact on the evolution error are then examined. This quantity varies from 20% to 100% by slice of 10% to see the evolution of the associated error, in order to see a stabilisation and then fix the size of the training base. For visualization purposes, the results is presented with the NSE error but the behaviour is the same for the MAE error (Fig. 7).

A clear improvement of the predictions and a stabilisation of the error can be observed when the training size reaches 50%, corresponding to data from April 1983 to February 1995. Over this period of time, two time intervals can be distinguished: the first one from 1983 to 1992 where high concentrations of ^{137}Cs are found in the SPM of Rhône-Vallabregues station which is the result of the fallout from Chernoby in l 1986 as well as an important period of release from Marcoule. The second period from 1992 to 1995 is characterised by a very strong decrease in ^{137}Cs concentrations due to the STEL facility installed in 1992 that improved the treatment of the nuclear wastes. It can be assumed that the post-1995 information, which is also characterised by lower concentrations does not provide more information on the behaviour of the system, which explains the stabilisation of the error.

5. Sensitivity analysis

Sensitivity analysis in neural networks with a complex architecture remains a delicate subject, notably because of the large number of hyperparameters and parameters of these models. It is still today a subject in full development (Finale Doshi-Velez, 2017). The Permutation feature importance, a simplistic approach is proposed to try to bring elements of answer on the importance of the various exogenous variables. Initially used for random forests, it is applicable to any model (Wei et al., 2015). This method has the advantage of not requiring a retraining phase for the model or long simulations, which can be costly in terms of computing time for HRHN. The concept is straightforward: we measure the importance of a feature by calculating the increase in the model's prediction error after permuting the feature. A feature (or a variable) is “important” if shuffling its values increases the model error, because in this case the model relied on the feature for the prediction. A feature is “unimportant” if shuffling its values leaves the model error unchanged, because in this case the model ignored the feature for the prediction. This method is applied to the test set rather than the training set to assess the importance of the variables in the ability of the model to generalise to the unknown data.If the training set is used to assess the importance of variables, this may lead to an overestimation of the importance of some variables, as the model has been optimised to minimise the error on these specific data. In contrast, the test set is used to assess the ability of the model to generalise to new data. By calculating the importance of the variables on the test set, we can assess the importance of the variables in the ability of the model to generalise and predict new data. The following Table 3 shows the results obtained. They correspond to the average of 100 simulations of Permutation feature importance for each variable.

This sensitivity analysis highlights the importance of two variables: the minimum water discharge and release. Initial tests with all the discharge water variables (min, max mean) showed a neutral or even negative influence of the max and mean discharge. The fluctuations of the three discharge data variables are related, so it is possible that the max and mean dischage do not provide additional information or even redundant information harmful to the algorithm as for debit-max with a

Table 1
Study of the NSE and MAE according to the different training bases. The red boxes show the results for the training base concerned.

Training base	Rhône-Vallabregues		Loire-Ouzouer		Loire-Muides	
	MAE	NSE	MAE	NSE	MAE	NSE
Rhône-Vallabregues	70	0.89	4.86	0.53	8	0.35
Loire-Ouzouer	143	0.37	3.9	0.66	9.2	0.16
Loire-Muides	180	0.2	5.5	0.45	8.5	0.25

Table 2
Values of the hyperameters selected following the optimisation process.

Hyperparameter	nbr-filters -cnn	dim-filters -cnn	dim-max -pooling	dim-RHN	nbr-layer -RHN	Length -sequence
Value	[64, 8]	[5, 3]	[2, 3]	64	4	8

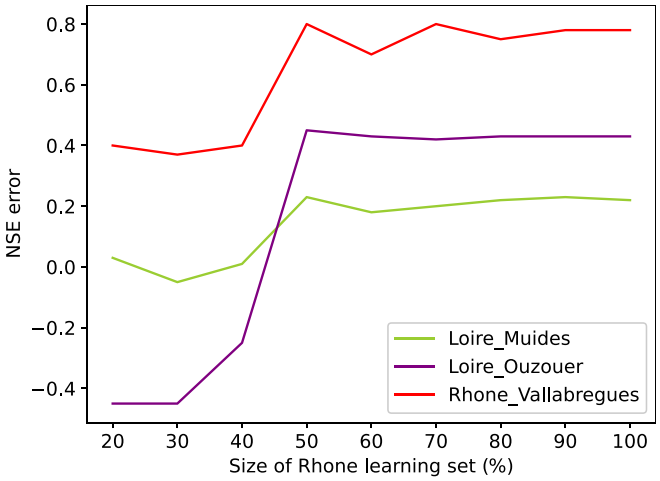


Fig. 7. Study of the evolution of the NSE error according to the size of the training base.

Table 3
Results of the sensitivity analysis: percentage importance of each variable.

Variables	Min water discharge	Nuclear release	Washout flux
Importance	0.456	0.370	0.172

negative score. It was therefore decided to remove these variables from our set of input variables. In addition, these initial tests showed a very weak influence (only 0.03) of the washout flux in relation to the discharge. This sensitivity was a little too close to the characteristics of the Rhône-Vallabregues station, which could indicate a tendency to overfitting. Indeed, on this site the releases of Marcoule were so large that they masked the rest of the contamination. The use of the rescale process provides a more independent sensitivity.

6. DA-RNN (Dual-Stage Attention-Based Recurrent Neural Network)

As in the article Tao et al. (2016), the performance of HRHN is compared with another encoder-decoder model, the DA-RNN (Dual-Stage Attention-Based Recurrent Neural Network) model, briefly mentioned in the introduction. The DA-RNN model is multi-attentive, there are two attention mechanisms: 'spatial attention' in the encoder and 'temporal attention' in the decoder. Each of the attention mechanisms is associated with a recurrent layer of the LSTM type Hochreiter and Schmidhuber (1997). The spatial attention mechanism processes the set of exogenous series and a spatial slice taken for an instant t (i.e. the set of values of the exogenous series taken at the same time). The objective is to weight the importance of the spatial slice in relation to the set of exogenous series. At each time, a weighted spatial slice is obtained, which is processed by an LSTM layer to recover a series of "t" hidden states. Subsequently, in the decoder, the temporal attention mechanism processes these t hidden states in order to associate a weighting on the importance between the different times for the calculation of a context vector transmitted to the LSTM layer for the final prediction.

Due to its structure, DA-RNN has much fewer hyperparameters than HRHN.The main hyperparameter is the dimension of the LSTM layers, the other hyperparameters are related to regularization (drop), to the dimensions of the data (batch-size and history length) and finally to the training plan. To determine these hyperparameters, the HyperBand algorithm is also used. Table 4 shows the hyperparameters obtained.

DA-RNN was trained under exactly the same conditions as HRHN: same number of epochs, choice of optimizer and same training base (Rhone-Vallabregues). The next section presents the results obtained with HRHN and the DA-RNN model.

7. Results

7.1. Reminder of comparison tools and metrics

To assess the quality of the predictions obtained, two metrics are added added in addition to MAE and NSE: RMSE (root-mean-square error) and bias:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}} \quad \text{et} \quad Bias = \frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)}{T}$$

As a reminder, the bias is used to assess whether or not the predictions are accurate and whether the model tends to over- or underestimate the values of the variable of interest. The lower the bias (close to 0), the better the prediction.

Taylor diagrams will also be used to analyze predictions. Taylor diagrams Taylor (2001) are used to graphically summarise the degree of correspondence between a model (or a set of models) and observations.

Table 4
Values of the hyperameters of DA-RNN following the optimisation process.

dim-LSTM	8
lenght-sequence	4

The similarity between two models is quantified in terms of the correlation, the root mean square difference (RMSD) and the amplitude of their variations (represented by their standard deviations). These diagrams are particularly useful for assessing the multiple aspects of complex models or for measuring the relative competence of different models. Simulated models that match observations well are located closest to the point marked 'observation' on the x-axis. Models that have relatively high correlation and low RMSD are considered to perform best.

7.2. HRHN and DA-RNN results

In this section, the graphics (Fig. 8) present the prediction obtained with the optimised HRHN and the optimised DA-RNN with the Rhône-Vallabregues as training base.

Observation-measurement representations can be used to quickly visualise the general behaviour of the prediction in relation to the $x = y$ curve.

For HRHN, the results of the predictions are in agreement with reality, the main trends are well reproduced. However, it can be seen that the peaks are well monitored but not always with the right amplitude: overestimation at the end of the Loire-Muides chronicle, underestimation of the 1990 peak at Loire-Ouzouer and underestimation of the peaks at the beginning of the chronicle at Rhone-Vallabrègues.

For the DA-RNN, the model cannot correctly minimise the error on the training set. The prediction on the training set is quite poor with permanent underestimation.On the system of Loire-Ouzouer, this is also a permanent overestimation. Finally on the system of Loire-Muides, the peak of 1990 has an amplitude closer to the measured values but is shifted with respect to the observations.

The Table 5 compares three metrics: RMSE, MAE and NSE and the bias. In general, HRHN performs better than DA-RNN. For both models, the system that is least well predicted is Loire-Muides. Both models are in difficulty because of the 1990 peak. The DA-RNN bias shows that DA-RNN underestimates predictions on the training set but overestimates predictions on the test sets. Generalization to different hydrological systems seems difficult for this model in particular because of the difficulties encountered during the training phase. For the HRHN model, there is a slight over-prediction of the training set with an under-prediction on the test sets. However, the bias for HRHN remains relatively low for all the systems.

We present Taylor diagrams (Fig. 9) for the two models HRNN and DA-RNN relative to the set of training observations, from the two test bases.The Table 6 summarises the results obtained.

For the Rhône-Vallabregues: the correlation between the HRHN models and the observations is stronger than for the DA-RNN model and the RMSD error is lower for HRHN. The standard deviation of the DA-RNN predictions is well below the standard deviation of the observations, indicated by the red arc at 442 Bq/kg. It is remarkable that the HRHN model provides the same standard deviation as the observations. The DA-RNN model has too little spatial variability compared with the observations, since its standard deviation is well below the standard deviation of the observations. The model closest to the observations would be the one with the lowest root mean square errors, i.e. the HRHN model.

For the Loire-Muides: the correlation between the HRHN models and the observations is once again stronger than for the DA-RNN model and the RMSD error is lower in the case of HRHN. The standard deviation of the DA-RNN predictions is well below the observed standard deviation, indicated by the red arc at 51Bq/kg. The HRHN model also has a standard deviation well below that of the observations. Despite this, we prefer the HRHN model, which has a lower RMSD and a high correlation with the observations.

For the Loire-Ouzouer: the correlation between the HRHN model and the observations is once again stronger than the DA-RNN model. In addition, the centred RMSD error is lower for the HRHN model. The

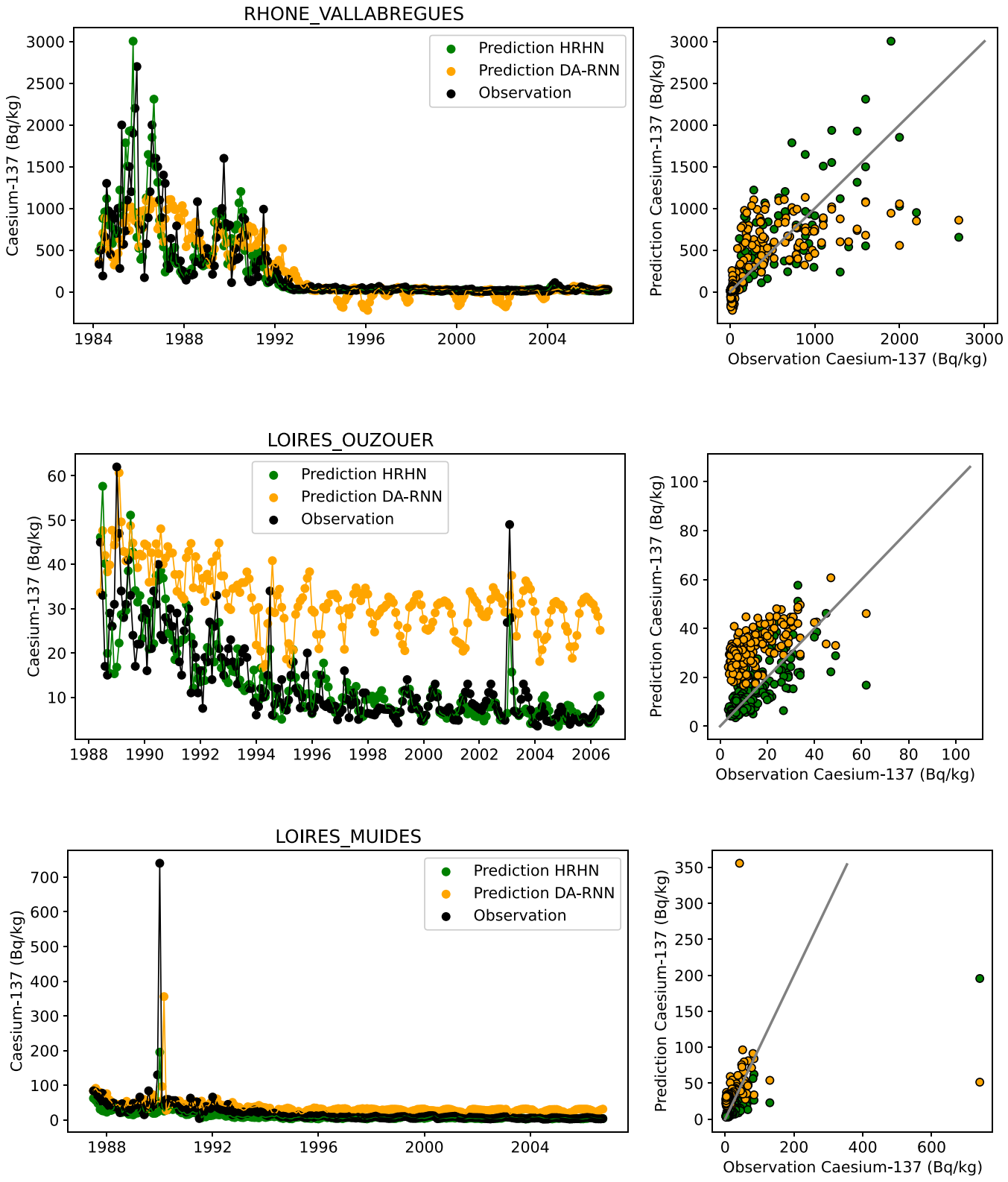


Fig. 8. For the three systems Rhone-Vallabregues, Loire-Ouzouer and Loire-Muides respectively: on the left, the time serie of the prediction of the concentrations of Cs137 (Bq/kg) in SPM, on the right, a comparison of observation and prediction.

standard deviation of the DA-RNN predictions is lower than the standard deviation of the observations, indicated by the red arc at 9.8 Bq/kg. It is remarkable that the HRHN model provides the same standard deviation as the observations. The DA-RNN model has too little spatial variability compared with the observations, since its standard deviation is less than the standard deviation of the observations. The model closest to the observations would be the one with the smallest squared errors, i.e. the HRHN model.

Table 5
Values of the metric and bias for HRHN and DA-RNN.

Modèle	Rhone-Vallabregues DA-RNN | HRHN	Loire-Ouzouer DA-RNN | HRHN	Loire-Muides DA-RNN | HRHN
RMSE	270|161	19|6.9	53|46
NSE	0.63|0.89	−3|0.6	−0.07|0.25
MAE	150|70	21|4	23|8.5
Bias	−29.1 |0.95	19.1|-0.068	14.9|-3.46

7.3. Discussion

The visualization of the results (Taylor diagram and analysis of the metrics) clearly show that HRHN is better suited to our study than DA-RNN. The appearance of negative values in the prediction of the DA-RNN model on the training set (Fig. 8, Rhone-Vallabregues) underlines the difficulty of the optimisation, even though it was carried out under exactly the same conditions as HRHN. This negative data may demonstrate that the DA-RNN model does not correctly capture the complex relationships present in the data. The main difference between the DA-RNN model and HRHN is the processing of model inputs (release, water discharge and washout flux) at the encoder level. In the DA-RNN model, there are a spatial attention mechanism and in HRNN a convolution layer. These two elements aim to analyze in a "spatial" manner (by association with a spatial cut for the attention mechanism or by filter for the convolution) the series of inputs in order to identify patterns and extract the input information. For the problem considered, given the results, the convolution layers seem more suitable for predicting caesium-137 concentrations. Better characterization of the input data allows better prediction of the target variable. Several advantages of convolution layers can be mentioned to explain this improvement:

- **Time translation invariance:** convolution layers are able to capture temporal patterns independently of their exact location in the sequence. This provides a degree of temporal translation invariance, which is essential for modelling sequential data.
- **Reducing temporal dimensionality:** The use of subsampling (pooling) with convolution layers reduces the temporal

Table 6
Values of the Correlation, RMSD and Standard deviation for HRHN and DA-RNN.

Modèle	Rhone-Vallabregues DA-RNN | HRHN	Loire-Ouzouer DA-RNN | HRHN	Loire-Muides DA-RNN | HRHN
Correlation	0.8|0.92	0.53|0.74	0.2|0.9
RMSD	260|171	8|7	51|37
Standard deviation	353|442	7|9.6	24|15

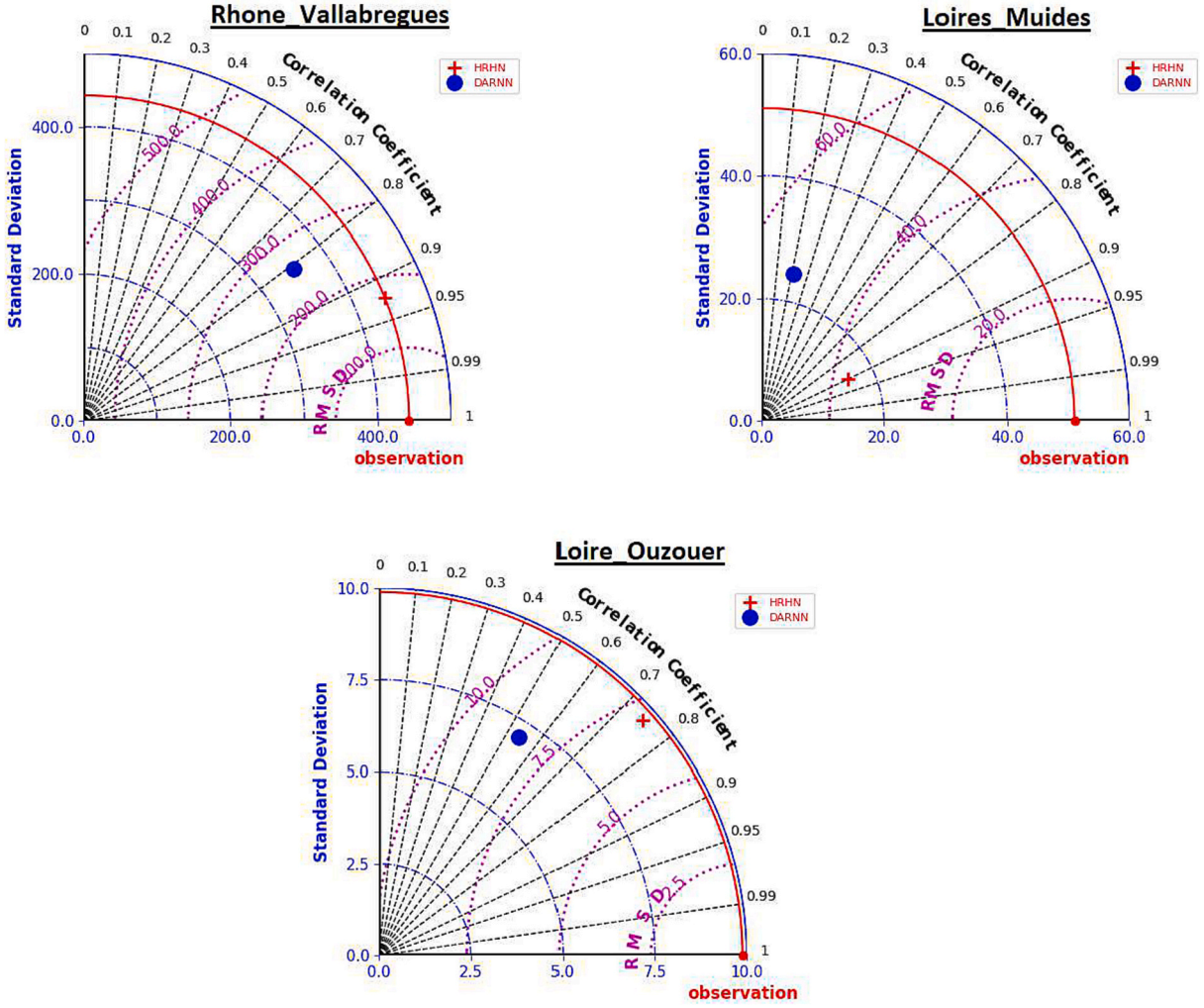


Fig. 9. Taylor diagram showing a statistical comparison with the observations of the two models HRHN and DA-RNN predictions on the training set (Rhône-Vallabregues) and on the two test sets (Loire-Muides and Loire-Ouzouer). The purple contours indicate the RMSD value. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

dimensionality of the data, which is useful for simplifying the model and avoiding overfitting.

- **Ability to learn sequential patterns:** By stacking multiple convolution layers, models can learn hierarchies of sequential patterns, where the first layers detect simple patterns and subsequent layers combine these patterns to detect more complex patterns.
- **Reuse of temporal patterns:** Temporal patterns learned by convolution filters in one part of the sequence can be reused in other parts of the sequence. This makes it easier to generalise the model to different parts of the sequence.

However, convolution can also present difficulties in capturing some features. Indeed, the poor detection of the 1990 peak at the Loires-Muides site on HRHN prediction could be explained by the convolution filter being unsuitable for detecting very spontaneous and chronically isolated phenomena. Indeed, we only find one event of this type in the chronicle of Loire-Muides.

8. Conclusion

In this article, a multivariate modelling by a neural network encoder decoder the Hierarchical Attention-Based Recurrent Highway Networks Tao et al. (2016)is proposed for the prediction of ^{137}Cs concentrations in SPM. This model extracts the maximum of information from past exogenous variable from river (water discharge, release and washout flux) with encoder part using convolution layer and recurrent layer (RHN) to generate a latent representation. A layer of Hierarchical Attention weighs the importance of this representation. Then the decoder part processes this representation with the past history of the target variable using an recurrent layer (RHN) for predict the future concentration of ^{137}Cs.

The model has been fully implemented on Python and the hyperparameters of the model were optimised using the custom HyperBand algorithm which allows to optimize the network architecture with the optimal length of the history.

Once optimised, the model generates first results in agreement with the real concentration curves by correctly following the main trends on different rivers.

The originality of this work lies in the fact that it is able to provide predictions for different hydrological system. In fact, the three explanatory variables retained (discharge data, release and washout flux) are the three major components influencing the concentration of ^{137}Cs whatever the river studied (Eyrolle et al., 2020b).So the HRHN model has demonstrated its robustness by being applicable to several rivers (Loire and Rhone) and several geographical sites with a limited number of variables. Another DA-RNN encoder-decoder architecture was implemented to confirm the relevance of the HRHN architecture. But DA-RNN's spatial attention mechanism is less efficient than the convolutional layers. Better performance was obtained with HRHN on all 3 systems. Finally, the sensitivity analysis should be improved to better capture the richness of the model. There are several ways of improving the model's predictions. The addition of covariates with information on tributaries could be an interesting possibility, as Lepage et al. (2023) has shown, the importance of information on tributaries, and more precisely the discharge. Furthermore, the calculation of the washout flux remains an estimate, and having values closer to the true values would probably improve the model.

CRediT authorship contribution statement

Kathleen Pele: Conceptualization, Data curation, Formal analysis, Methodology, Resources, Software, Visualization, Writing – original draft, Writing – review & editing. **Valérie Nicoulaud-Gouin:** Methodology, Project administration, Validation, Visualization, Writing – review & editing. **Hugo Lepage:** Project administration, Supervision, Validation, Visualization, Writing – review & editing.

Declaration of competing interest

None.

Data availability

All methods were written in the Python language and in the PyCharm environment (https://www.jetbrains.com/pycharm/). machine learning methods were developed using the freely available Keras library (Chollet, 2018) at https://github.com/fchollet/keras with Tensorflow backend. Free and open-source codes have been employed to perform the overall analysis: the hyperparameters ofthe model were optimised using the custom HyperBand algorithm which allows to optimize the network architecture with the optimal length of the history.

Acknowledgement

The authors would like to thank the consortium ANR “Trajectoire” (ANR-19-CE3- 0009, 2020–2024) especially F. Eyrolle and D. Claval and the ARCENV data processing team especially O.Pierrard.

References

Aggarwal, C.C., 2018. Neural Networks and Deep Learning.
Bahdanau, D., Cho, K., Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. ArXiv 1409.
Borzilov, V., Konoplev, A., Revina, S., Bobovnikova, T.I., Lyutik, P., Shveikin, Y.V., Shcherbak, A., 1988. Experimental investigation of washout of radionuclides deposited on soil as a result of the Chernobyl nuclear power plant accident. Sov. Meteorol. Hydrol. 11, 43–53.
Chen, B., Li, W., 2020. Multitime resolution hierarchical attention-based recurrent highway networks for taxi demand prediction. Math. Probl. Eng. 2020, 1–10. https://doi.org/10.1155/2020/4173094.
Cho, K., Merrienboer, B.V., Bahdanau, D., Bengio, Y., 2014a. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. https://doi.org/10.48550/arXiv.1409.1259.
Cho, K., Merrienboer, B.V., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014b. Learning Phrase Representations Using rnn Encoder-Decoder for Statistical Machine Translation. https://doi.org/10.48550/arXiv.1406.1078.
Chollet, F., 2018. Deep Learning with Python. Manning. URL: https://tanthiamhuat.files.wordpress.com/2018/03/deeplearningwithpython.pdf.
Delile, H., Masson, M., Miège, C., Le Coz, J., Poulier, G., Le Bescond, C., Radakovitch, O., Coquery, M., 2020. Hydro-climatic drivers of land-based organic and inorganic particulate micropollutant fluxes: the regime of the largest river water inflow of the mediterranean sea. Water Res. 185.
Delmas, M., Garcia-Sanchez, L., Nicoulaud-Gouin, V., Onda, Y., 2017. Improving transfer functions to describe radiocesium wash-off fluxes for the niida river by a bayesian approach. J. Environ. Radioact. 167, 100–109. https://doi.org/10.1016/j.jenvrad.2016.11.002.
Dragović, S., 2022. Artificial neural network modeling in environmental radioactivity studies – a review. Sci. Total Environ. 847, 157526.
Eyrolle, F., Lepage, H., Antonelli, C., Morereau, A., Cossonnet, C., Boyer, P., Gurriaran, R., 2020a. Radionuclides in waters and suspended sediments in the rhone river (France) - current contents, anthropic pressures and trajectories. Sci. Total Environ. 723, 137873 https://doi.org/10.1016/j.scitotenv.2020.137873.
Eyrolle, F., Lepage, H., Antonelli, C., Morereau, A., Cossonnet, C., Boyer, P., Gurriaran, R., 2020b. Radionuclides in waters and suspended sediments in the rhone river (France) current contents, anthropic pressures and trajectories. Sci. Total Environ. 723, 137873.
Fawaz, H.I., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A., 2019. Deep learning for time series classification: a review. Data Min. Knowl. Disc. 33, 917–963. https://doi.org/10.1007/s10618-019-00619-1.
Finale Doshi-Velez, B.K., 2017. Towards a Rigorous Science of Interpretable Machine Learning. arXiv. https://doi.org/10.48550/arXiv.1702.08608.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press. URL: http:/www.deeplearningbook.org.
Goutal, N., Luck, M., Boyer, P., Monte, L., Siclet, F., Angeli, G., 2008. Assessment, validation and intercomparison of operational models for predicting tritium migration from routine discharges of nuclear power plants: the case of loire river. J. Environ. Radioact. 99, 367–382. https://doi.org/10.1016/j.jenvrad.2007.10.016.
Goyal, M., 2014. Modeling of sediment yield prediction using m5 model tree algorithm and wavelet regression. Water Resour. Manag. 28, 1991–2003. https://doi.org/10.1007/s11269-014-0590-6.
Hirose, K., Povinec, P., 2022. Ten years of investigations of Fukushima radionuclides in the environment: a review on process studies in environmental compartments. J. Environ. Radioact. 251–252 https://doi.org/10.1016/j.jenvrad.2022.106929.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.
Ikenoue, T., Shimadera, H., Nakanishi, T., Kondo, A., 2023. Thirty-year simulation of environmental fate of 137cs in the abukuma river basin considering the characteristics of 137cs behavior in land uses. Sci. Total Environ. 876.
Iwasaki, T., Nabi, M., Shimizu, Y., Kimura, I., 2015. Computational modeling of 137cs contaminant transfer associated with sediment transport in abukuma river. J. Environ. Radioact. 139, 416–426.
Kashparov, V., Salbu, B., Levchuk, S., Protsak, V., Maloshtan, I., Simonucci, C., Courbet, C., Nguyen, H., Sanzharova, N., Zabrotsky, V., 2019. Environmental behaviour of radioactive particles from Chernobyl. J. Environ. Radioact. 208–209 https://doi.org/10.1016/j.jenvrad.2019.106025.
Khanbilvardi, R., Shestopalov, V., Onishchenko, I., Bublyas, V., Gudzenko, V., 1999. Role of erosion process in transfer of radionuclides: result of field experiments1. JAWRA J. Am. Water Res. Assoc. 35, 887–898. https://doi.org/10.1111/j.1752-1688.1999.tb04182.x.
Konoplev, A., Kanivets, V., Laptev, G., Voitsekhovich, O., Zhukova, O., Germenchuk, M., 2020. Long-term dynamics of the Chernobyl-derived radionuclides in rivers and lakes. Behav. Radionuclides Environ. II, 323–348.
Kryshev, I., 1995. Radioactive contamination of aquatic ecosystems following the Chernobyl accident. J. Environ. Radioact. 27, 207–219.
Kulahci, F., Özer, A., Doğru, M., 2006. Prediction of the radioactivity in hazar lake (sivrice, Turkey) by artificial neural networks. J. Radioanal. Nucl. Chem. 269, 63–68. https://doi.org/10.1007/s10967-006-0230-6.
Lecun, Y., Bengio, Y., 1995. Convolutional Networks for Images, Speech, and Time-Series. MIT Press.
Lepage, H., Laceby, J., Bonté, P., Joron, J., Onda, Y., Lefèvre, I., Ayrault, S., Evrard, O., 2016. Investigating the source of radiocesium contaminated sediment in two Fukushima coastal catchments with sediment tracing techniques. Anthropocene 13.
Lepage, H., Nicoulaud-Gouin, V., Pele, K., Boyer, P., 2023. Use of machine learning and deep learning to predict particulate 137cs concentrations in a nuclearized river. J. Environ. Radioact. 270 https://doi.org/10.1016/j.jenvrad.2023.107294.
Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., Talwalkar, A., 2018. Hyperband: a novel bandit-based approach to hyperparameter optimization. J. Mach. Learn. Res. 18, 1–52.
Meusburger, K., Evrard, O., Alewell, C., Borrelli, P., Cinelli, G., Ketterer, M., Mabit, L., Panagos, P., van Oost, K., Ballabio, C., 2020. Plutonium Aided Reconstruction of Caesium Atmospheric Fallout in European Topsoils. Scientific Reports.
Moatar, F., Dupont, N., 2016. La Loire fluviale et estuarienne: un milieu en évolution.
Moatar, F., Fessant, F., Poirel, A., 1999. ph modelling by neural networks. Application of control and validation data series in the middle loire river. Ecol. Model. 120, 141–156. URL: https://www.sciencedirect.com/science/article/pii/S0304380099000988.
Oludare Isaac, A., Aman, J., Abiodun Esther, O., Kemi, V.D., Nachaat AbdElatif, M., Humaira, A., 2018. State-of-the-art in artificial neural network applications: a survey. Heliyon 4. https://doi.org/10.1016/j.heliyon.2018.e00938.
Poulier, G., Launay, M., Le Bescond, C., Thollet, F., Coquery, M., Le Coz, J., 2019. Combining flux monitoring and data reconstruction to establish annual budgets of suspended particulate matter, mercury and pcb in the rhône river from Lake Geneva to the mediterranean sea. Sci. Total Environ. 658, 457–473.
Qin, Y., Song, D., Cheng, H., Cheng, W., Jiang, G., Cottrell, G., 2017. A dual-stage attention-based recurrent neural network for time series prediction. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence.
Rajaee, T., Khani, S., Ravansalar, M., 2020. Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: a review. Chemom. Intell. Lab. Syst. 200, 103978.
Roussel-Debel, S., Renaud, P., Métivier, J.M., 2007. 137cs in french soils: deposition patterns and 15-year evolution. Sci. Total Environ. 374, 388–398. https://doi.org/10.1016/j.scitotenv.2006.12.037.
Schmidhuber, J., 1992. Learning complex, extended sequences using the principle of history compression. Neural Comput. 4, 234–242. https://doi.org/10.1162/neco.1992.4.2.234 doi: 10.1162/neco.1992.4.2.234, arXiv:https://direct.mit.edu/neco/article-pdf/4/2/234/812272/neco.1992.4.2.234.pdf.
Shoham, R., Permuter, H., 2018. Highway State Gating for Recurrent Highway Networks: Improving Information Flow through Time. arXiv:1805.09238.
Shuryak, I., 2022. Machine learning analysis of 137cs contamination of terrestrial plants after the Fukushima accident using the random forest algorithm. J. Environ. Radioact. 241, 106772 https://doi.org/10.1016/j.jenvrad.2021.106772.
Stamenković, L., Mrazovac Kurilic, S., Ulniković, V., 2020. Prediction of nitrate concentration in Danube river water by using artificial neural networks. Water Supply 20, 2119–2132. https://doi.org/10.2166/ws.2020.104.
Takahashi, Y., Fan, Q., Suga, H., Tanaka, K., Sakaguchi, A., Takeichi, Y., Ono, K., Mase, K., Kato, K., Kanivets, V.V., 2017. Comparison of solid-water partitions of radiocesium in river waters in Fukushima and Chernobyl areas. Sci. Rep. 7, 1–11.
Tao, Y., Ma, L., Zhang, W., 2016. Hierarchical attention-based recurrent highway networks for time series prediction. IEEE Trans. Intell. Transp. Syst. 17, 2479–2489.
Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. Atmos. 106, 7183–7192. https://doi.org/10.1029/2000JD900719.
Tiwari, S.K., Babbar, R., Kaur, G., 2018. Performance evaluation of two anfis models for predicting water quality index of river satluj (India). Adv. Civil Eng. 2018, 1–10.
Tomczak, W., Boyer, P., Eyrolle, F., Radakovitch, O., Krimissa, M., Lepage, H., Amielh, M., Anselmet, F., 2021. Modelling of solid/liquid fractionation of trace metals for suspended sediments according to the hydro-sedimentary conditions of rivers-application to 137cs in the rhône river (France). Environ. Model. Softw. 145.
Tracy, B., Carini, F., Barabash, S., Berkovskyy, V., Brittain, J., Chouhan, S., Eleftheriou, G., Iosjpe, M., Monte, L., Psaltaki, M., Shen, J., Tschiersch, J., Turcanu, C., 2013. The sensitivity of different environments to radioactive contamination. J. Environ. Radioact. 122, 1–8. https://doi.org/10.1016/j.jenvrad.2013.02.015.
Vrel, A., 2012. Reconstitution de l'historique des apports en radionucléides et contaminants métalliques à l'estuaire fluvial de la seine par l'analyse de leur enregistrement sédimentaire.
Wei, P., Zhenzhou, L., Song, J., 2015. Variable importance analysis: a comprehensive review. Reliab. Eng. Syst. Saf. 142, 399–432. https://doi.org/10.1016/j.ress.2015.05.018.
Yang, D., Li, S., Peng, Z., Wang, P., Wang, J., Yang, H., 2019. Mf-cnn: traffic flow prediction using convolutional neural network and multi-features fusion. IEICE Trans. Inf. Syst. E102.D, 1526–1536. https://doi.org/10.1587/transinf.2018EDP7330.
Yaseen, Z.M., 2021. An insight into machine learning models era in simulating soil, water bodies and adsorption heavy metals: review, challenges and solutions. Chemosphere 277, 130126. URL: https://www.sciencedirect.com/science/article/pii/S0045653521005956.
Yeo, I., Johnson, R., 2000. A new family of power transformations to improve normality or symmetry. Biometrika 87, 954–959.
Yoshimura, K., Onda, Y., Sakaguchi, A., Yamamoto, M., Matsuura, Y., 2015. An extensive study of the concentrations of particulate/dissolved radiocaesium derived from the Fukushima dai-ichi nuclear power plant accident in various river systems and their relationship with catchment inventory. J. Environ. Radioact. 139, 370–378.
Zhang, C., Nguyen, T., Sah, S., Ptucha, R., Loui, A., Salvaggio, C., 2017. Batch-normalized recurrent highway networks. In: 2017 IEEE International Conference on Image Processing (ICIP). IEEE. https://doi.org/10.1109/icip.2017.8296359. URL: doi:10.1109%2Ficip.2017.8296359.
Zilly, J.G., Srivastava, R.K., Koutn'k, J., Schmidhuber, J., 2016. Recurrent Highway Networks. arXiv e-Prints.