

Comment prévoir plus précisément les indisponibilités climatiques des réacteurs nucléaires ?

Le défi

Plus des deux-tiers de l'électricité produite en France est d'origine nucléaire. Depuis plusieurs années, cette production est régulièrement perturbée lors de vagues de chaleur ou de sécheresses. Ces indisponibilités d'origine climatique peuvent avoir des conséquences importantes, notamment sur les marchés de l'électricité : la canicule de juillet 2019, par exemple, a forcé jusqu'à 9 réacteurs à réduire ou baisser simultanément leurs productions, ce qui a entraîné une hausse du prix spot de l'électricité au-dessus de 70€/MWh, son plus haut niveau en 2019 hors période hivernale.

Le problème

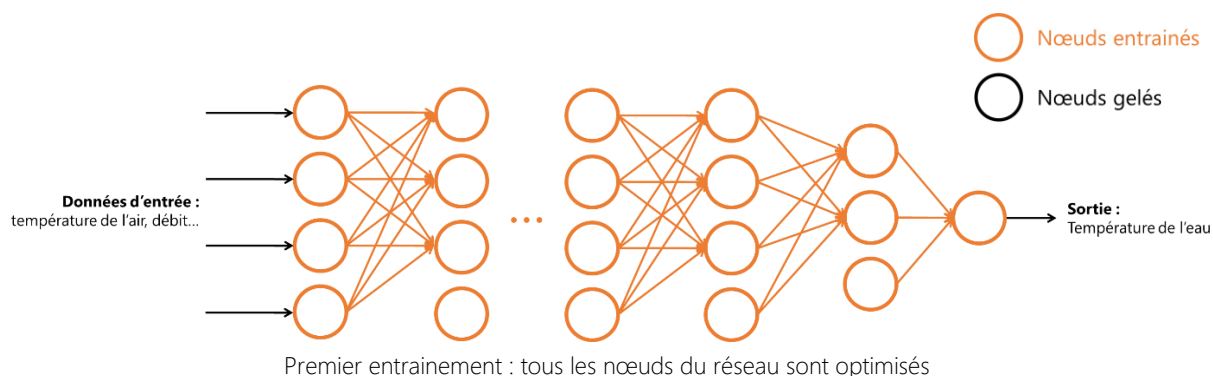
L'un des principaux déclencheurs de ces indisponibilités est la température des fleuves utilisés pour le refroidissement des réacteurs. Lorsque la température en amont est trop élevée, les centrales ne peuvent plus refroidir leurs turbines en respectant les limites de température en aval et sont contraintes de réduire leur puissance. Des données précises sur la température de l'eau sont donc essentielles pour anticiper la disponibilité des réacteurs nucléaires en période de chaleur ou de sécheresse.

Cependant, il est difficile de créer des modèles de température de l'eau performants car les données publiques disponibles diffèrent sensiblement des mesures d'EDF. Par conséquent, un modèle entraîné ou calibré sur ces données biaisées ne peut pas prédire de façon fiable les décisions d'arrêt ou de réduction de charge de l'opérateur.

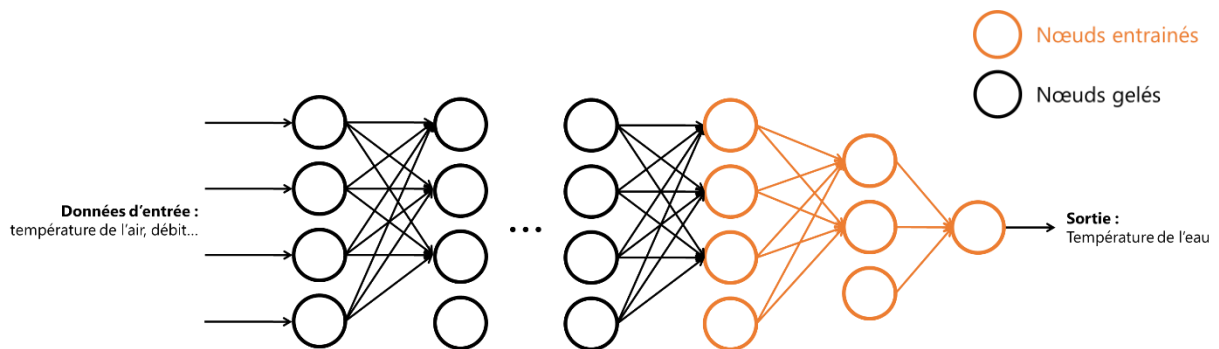
Notre solution

Callendar a mis au point une solution innovante pour prédire précisément la température de l'eau des fleuves au niveau des centrales nucléaires françaises :

1. Tout d'abord, nous avons **entraîné un réseau de neurones profond sur des mesures publiques** de température de l'eau. L'objectif de ce premier entraînement est d'apprendre la physique générale de la rivière : comment la température de l'eau réagit-elle aux variations de la température de l'air, du débit de la rivière, etc.



2. Nous utilisons ensuite des techniques de datamining pour **collecter des mesures de température de l'eau réalisées par l'exploitant**. EDF ne publie pas régulièrement ces données mais en analysant diverses sources (rapports environnementaux, newsletter, communiqué de presse...), nous pouvons créer un petit jeu de données de température non biaisé.
3. Ce jeu de données est trop limité pour entraîner un modèle complexe, nous ne l'utilisons que pour **réentraîner partiellement le réseau de neurone existant à partir de données non-biaisées**. Avec cette technique, appelée « apprentissage par transfert », le réseau garde en mémoire ce qu'il a appris des données publiques pendant que ses dernières couches sont ajustées pour corriger leurs biais.



Apprentissage par transfert : le réseau de neurones est réentraîné avec des données non biaisées, les premières couches sont gelées, seules les dernière couches sont optimisées

Un exemple

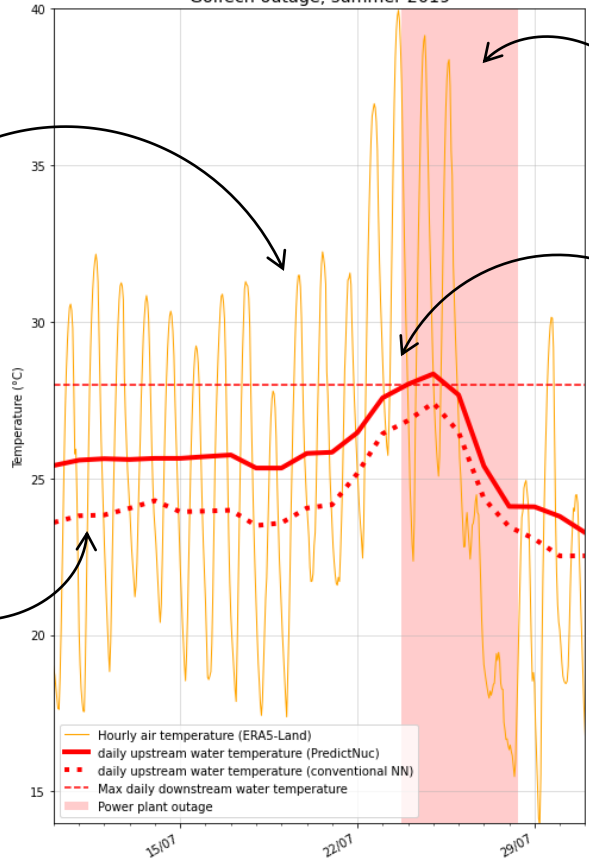
Cette approche permet de reproduire les indisponibilités passées du parc nucléaire avec des performances nettement améliorées, à la fois en termes de faux positifs et de faux négatifs.

Par exemple pour l'arrêt de Golfech pendant la vague de chaleur de juillet 2019 :

Golfech outage, summer 2019

A partir du 20 juillet la température de l'air augmente rapidement autour de Golfech.

Un modèle de machine learning entraîné à partir de données publiques donne des tendances similaires mais ne détecte pas le risque d'indisponibilité.



La vague de chaleur se traduit par une indisponibilité totale entre le 23 et le 28 juillet.

Notre modèle prédit correctement que la température de l'eau atteint 28°C, le maximum autorisé, le 23 juillet, entraînant l'arrêt de la centrale.

