

Atelier IRN-RHYMA : Les Outils d'Intelligence Artificielle En Géosciences Théories Et Exemples Pratiques En Hydrologie Et Érosion Des Sols

1. Présentation de l'atelier

Cet atelier s'inscrit dans le dispositif de recherche de l'IRD IRN RHYMA (International Research Network : Risques HYdrologiques au MAGhreb – Crues et Érosion des Sols). L'objectif central de l'IRN RHYMA-CES est de fédérer un collectif scientifique sur l'étude des risques de crues et d'érosion des sols au Maghreb, afin de partager et développer des moyens, des savoirs et des expertises autour de trois axes : améliorer la modélisation et la prévision de ces phénomènes pour répondre aux enjeux de suivi et de gestion de crise, prédire leur évolution dans des conditions futures probables pour soutenir la planification, et explorer des stratégies d'atténuation adaptées aux conditions locales pour renforcer la prévention.

Dans ce cadre, l'atelier vise à initier un public varié (professionnels, doctorants, étudiants) aux applications du Machine Learning et du Deep Learning pour l'étude des risques d'érosion et de crues. La première partie sera consacrée à la cartographie de la susceptibilité et de la vulnérabilité à l'érosion des sols, en présentant une démarche simple et accessible pour comprendre les bases méthodologiques, les types de données mobilisées et l'interprétation des résultats. L'objectif est de donner aux participants les outils pour identifier et représenter spatialement les zones les plus susceptibles de subir une érosion. La seconde partie sera dédiée à la prévision des séries temporelles liées aux crues, avec une introduction au Deep Learning et plus particulièrement aux réseaux LSTM (Long Short Term Memory). Appliquée aux chroniques de débit de ruissellement, cette formation permettra aux participants de découvrir les étapes clés d'un projet de modélisation temporelle et de mettre en pratique un premier modèle de prévision.

En combinant ces deux volets, spatial et temporel, l'atelier offre une formation intégrée qui contribue aux objectifs de l'IRN RHYMA : améliorer la compréhension, la représentation et l'anticipation des processus de crues et d'érosion des sols dans les contextes vulnérables du Maghreb.

2. Objectifs

- Favoriser le transfert de connaissances en proposant une formation accessible et progressive, permettant aux professionnels, doctorants et étudiants d'acquérir les bases nécessaires pour appliquer le Machine Learning et le Deep Learning aux problématiques de crues et d'érosion des sols.
- Offrir une formation pratique, efficace et reproductible, qui puisse être facilement adaptée et appliquée à d'autres contextes géographiques et environnementaux.
- Comprendre les notions fondamentales de susceptibilité et de vulnérabilité à l'érosion, ainsi que les enjeux liés à la modélisation des crues.
- Découvrir les bases du Machine Learning appliqué à la cartographie prédictive, en apprenant la démarche méthodologique pour construire une carte de susceptibilité et en identifiant les types de données mobilisables.
- Approfondir les bases du Deep Learning appliqué aux séries temporelles, en particulier les réseaux LSTM (Long Short Term Memory), afin de comprendre les étapes clés d'un projet de prévision des chroniques hydrologiques.
- Développer la capacité à interpréter et analyser les résultats produits par ces approches, qu'il s'agisse de matrices de confusion, de cartes de probabilité ou de sorties de modèles LSTM.

3. Contenu de l'atelier

3.1. Thème 1 : Application du Machine Learning pour la cartographie de la susceptibilité à l'érosion des sols (Durée : 3 sessions sur 3 jours)

3.1.1. Objectifs

- Comprendre les notions de *susceptibilité* à l'érosion.
- Acquérir une vue d'ensemble des méthodes de cartographie par Machine Learning.
- Être capable d'identifier les types de données nécessaires.
- Découvrir le fonctionnement d'un modèle ML simple (Random Forest).
- Interpréter les résultats : cartes, matrice de confusion, importance des variables.

3.1.2. Session 1 – Introduction (2h)

Objectif : poser le cadre général, vulgariser les concepts clés :

1. Enjeux liés à l'érosion des sols (exemples concrets)
2. Définitions : érosion, susceptibilité

3. Principes de la cartographie spatiale prédictive
4. Introduction au Machine Learning :
 - a. Définition ?
 - b. Pourquoi l'utiliser en cartographie environnementale ?
 - c. Exemples d'applications

3.1.3. Session 2 – Démarche théorique (3h)

Objectif : expliquer les étapes clés d'un projet de cartographie par ML.

1. Démarche méthodologique pas à pas :
 - a. Définir le problème et le type de risque
 - b. Collecte et préparation des données (type raster/vectoriel, résolution, variables explicatives : pente, pluie, occupation du sol, etc.)
 - c. Constitution des échantillons (points d'occurrence/absence)
 - d. Introduction au modèle Random Forest (principe de base, pourquoi il est adapté pour débiter)
 - e. Division en données d'entraînement/test
 - f. Génération de la carte de prédiction
2. Explication simplifiée des sorties du modèle :
 - a. Matrice de confusion
 - b. Interprétation d'une carte de probabilité de risque
 - c. Lecture des « *best input features* » (sans entrer dans le détail du *feature importance*)

3.1.4. Session 3 – Session pratique (3h à 4h)

Objectif : Application

1. Prise en main d'un notebook Python
2. Chargement des données : MNT, pluie, occupation de sol, etc.
3. Sélection des points d'échantillonnage
4. Entraînement d'un modèle Random Forest
5. Création de la carte de susceptibilité
6. Validation avec matrice de confusion
7. Validation avec des observation collecter sur Terrain.
8. Discussion et échange.

3.2. Thème 2 : Introduction à l'utilisation du Deep Learning pour la prédiction des séries temporelles en hydrologie.

Durée : 3 sessions (sur 3 jours)

3.2.1 Objectifs:

- Identifier les enjeux de la modélisation des séries temporelles.
- Acquérir une vue d'ensemble des approches du Deep Learning.
- Comprendre la logique d'un modèle Long Short Term Memory (LSTM) appliqué aux séries temporelles.
- Identifier les étapes clés d'un projet de prévision des chroniques du débit par LSTM.
- Être capable de construire un modèle LSTM pour la prédiction des séries chronologiques.

3.2.2. Session 1 – Introduction (2h)

Objectif :

1. Enjeux liés à la prévision des séries temporelles.
2. Pourquoi utiliser l'intelligence artificielle dans ce contexte ?
3. Rappel des séries temporelles :
 - a) Définition, propriétés (tendance, saisonnalité, bruit).
 - b) Problèmes fréquents dans les données hydrologiques (données manquantes, non-linéarité).
4. Présentation du Deep Learning :
 - a) Qu'est-ce qu'un réseau de neurones ?
 - b) Pourquoi les LSTM sont adaptés aux chroniques ?
 - c) Exemples d'applications dans l'hydrologie.

3.2.2. Session 2 – Démarche théorique (3h)

Objectif : Expliquer les étapes clés d'un projet de prévision par LSTM

1. Prétraitement des données :
 - a) Choix des variables d'entrée.
 - b) Standardisation des données.
2. Structure d'un modèle LSTM (simplifiée) :
 - a) Définition du nombre de couches.

- b) Fonctionnement des fenêtres de prévision.
 - c) Définition des hyperparamètres (nombre de neurones, batch size, epochs, dropout).
 - d) Importance de la division des jeux de données en entraînement-test.
 - e) Notion d'optimisation des hyperparamètres.
3. Évaluation des résultats :
- a) Courbes de prédiction vs observation.
 - b) Métriques de validation : R2, RMSE, NSE.

3.2.2. Session 3 – Session pratique (3h à 4h)

Objectif :

1. Création du fichier des données (CSV de chroniques hydrologiques).
2. Prise en main d'un notebook Python (Google Colab).
3. Chargement des données.
4. Explication du code et des paramètres modifiables.
5. Définition de la fenêtre de prévision.
6. Structuration du modèle (nombre couche, units, batch size, epochs...).
7. Lancement du modèle (entraînement et validation)
8. Visualisation des résultats (courbes, erreurs)
9. Modification des paramètres (optimisation manuelle) et relancement du code.
10. Lecture et comparaison des résultats avec discussion collective.