Analyse des données de pluies au Japon

Mathias Ballot

Mars 2024

Introduction

Le Japon est régulièrement touché par des désastres naturels tel que des typhons, des averses violentes ou des inondations. Le taux de précipitations par ans est deux fois plus élevé que la moyenne mondiale, et tombe en majorité durant la saison des pluies entre mai et août. Face a de tels évènements, le développement d'outils de mesure et de prédiction des précipitations urge.

Le Japon étant une île montagneuse l'eau des précipitations a tendance à rapidement ruisseler vers la mer. Des barrages permettent de contrôler le flux des précipitations, mais sont souvent débordées par les averses violentes devenues de plus en plus fréquentes [4]. Des plans d'actions prévoient le relâchement anticipée de l'eau stockée dans les barrages afin d'éviter le débordements lors de l'averse. Les barrages régulent à eux seules près de 50% de l'eau utiliser dans les foyers japonnais. L'énergie hydroélectrique qui représente 8% de la production électrique japonaise reste l'énergie renouvelable la plus fiable dans un tel pays. Afin d'optimiser la production électrique du barrage l'eau relâchée doit être remplacée par celle provenant des averse. Pour cela des estimations précises de la quantité de pluies tombant dans les alentours du barrage est nécessaire. Les prévisions doivent êtres estimées 2 à 3 heures en avances, ce qui correspond à l'afflux de l'eau des précipitations dans les réservoirs [5].

Pour répondre à ces besoins le Japon développe depuis les années 1950 des radars d'observations météorologique de plus en plus performants [1]. Ces radars qui n'étaient capable que de détecter des zones de pluies peuvent maintenant estimer l'intensité et la forme des gouttes de pluie. En 1964, un radar a été construit au point culminant du Japon : le mont Fuji. Il était à l'époque le radar le plus haut du monde.
Une autre manière de mesurer les précipitations consiste à disséminer des jauges de mesures sur le sol Japonais. Ces jauges permettent de mesurer avec précision la quantité de pluie en un point.

Les résultats suivants sont une analyse des données de pluies au Japon. Les modèles de machine développés à partir de ces données ne sont pas présenter dans ce papier, mais le code est accessible depuis le gitHub.

De tel prédictions servent à réguler les quantités d'eau stockée dans les barrages, à la préventions de désastres naturels, l'approvisionnement d'eau dans les foyers, l'agriculture...

Chapter 1

Analyse des données

1.1 Japan Meteorological Agency

La "Japan Meteorological Agency" décompose l'analyse des précipitations en 2 catégories [3].

- Estimation météorologique : Consiste à déterminer la quantité et l'intensité des précipitions passées ou présentes. Ces donnes sont collectées par des jauges et des radars.
- Prévision météorologique : prédictions des futures précipitations à partir d'un modèle météorologique. Ce modèle peut être basé sur des équations mathématique, des règles thermodynamiques, la théorie des fluides...

1.1.1 Estimation météorologique

Dans cette partie, nous nous intéressons aux estimations météorologiques. Pour estimer la pluie toute les 30 minutes avec une résolution spatiale de 1 km, la "Japan Meteorological Agency" utilise deux instruments de mesures :

- Les jauges de pluie : Conteneur physique collectant une certaine quantité de pluie. Les jauges sont des instruments de mesures fiables et précis. Une jauge ne peut mesure la pluie qu'en un seul emplacement, il est donc nécessaire d'en déployer partout afin de mesurer les précipitations sur le sol japonnais. Plus de 10.000 jauge sont disséminer a intervalle régulier de 7 km à travers le territoire japonnais, collectant des données toute les 10 minutes ou 1 heures.
- Les radars de pluie : Radar émettant des impulsions courtes de manières périodique. Les ondes émises sont partiellement réfléchies par les goûtes de pluies, donnant des informations sur l'intensité, la position et le déplacements des précipitations. Un radar peut couvrir a lui seule une surface de 500km^2. 46 radars mesurent les précipitations au Japon, générant une

cartographie des pluies avec une résolution spatiale de 1 km et à intervalle de 5 minutes.

Pour réaliser les estimations météorologique, les données radars sont sommées puis calibrer grâce aux mesures de jauges. L'atténuation des ondes radar et le caractère uniforme des précipitations sont utilisés pour affiner les résultats. Dans la figure 1.1 la figure centrale est obtenu en utilisant les mesures radars (à gauche) et les mesures de jauges (à droite). On remarque que les mesures de jauges sont ponctuels, contrairement aux mesures radars.

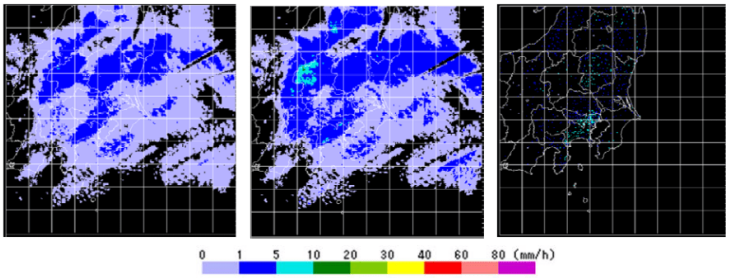


Figure 1.1: Gauche : mesures radar; Droite : mesures jauge; Centre : Estimations combinées

1.1.2 Prévisions météorologique

La prévision météorologique sur court terme (6 heures) se base sur 2 éléments :

- Extrapolation des estimations : Les estimations mesurées par les radars et jauge sont généralisée aux 6 heures suivantes. L'intensité et la direction du vent est pris en compte afin de prédire les mouvements des précipitations sur les prochaines heures. La topologie montagneuse du Japon causant les nuages à se condenser est aussi prise en compte.

- Modèle méso-échelle : Le terme méso-échelle est une échelle de grandeur représentant les phénomènes météorologique entre 2km et 2.000km (échelle planétaire). Un tel modèle se base sur des équations de masse, de dynamisme, d'énergie interne et toute ce qui touche la mécanique des fluides.

Les extrapolations sont précises sur les 3 premières heures de prévisions. Le modèle méso-échelle vient compléter les extrapolation sur des heures de prévision plus lointaines. Dans la figure 1.2 les extrapolations révèlent une zone intense de pluie au nord-est du Japon (encadré en rouge), alors que le modèle méso-échelle montre de forte précipitations au coeur du Japon (encadré en bleue). Le mélange des deux prévisions se rapproche des quantités de précipitation réellement obtenus.

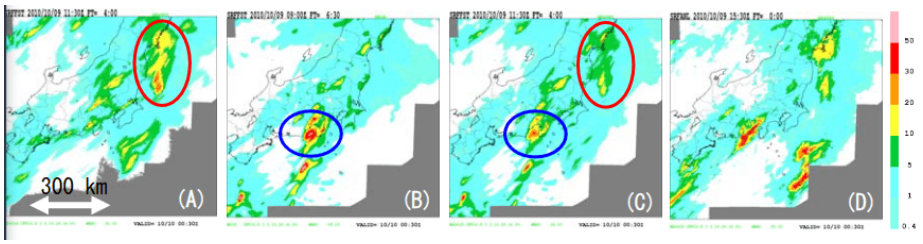


Figure 1.2: A: Extrapolation; B: Modèle méso-échelle; C : Combinaisons de A et B; D: Valeur réellement mesurée

1.1.3 Exemple d'application

Les prévisions météorologique sont utilisées pour mesurer l'indice d'humidité dans le sol. Des risques de glissement de terrains sont fréquent sur les sols ayant un fort indice. L'eau des précipitions restent stocke dans le sol pendant un certain temps avant de s'écouler, ces désastres peuvent donc être causé par des pluies ayant eu lieux quelques jours plus tôt. La figure 1.3 montre que le taux d'humidité dans le sol permet de prévoir un tel désastre.

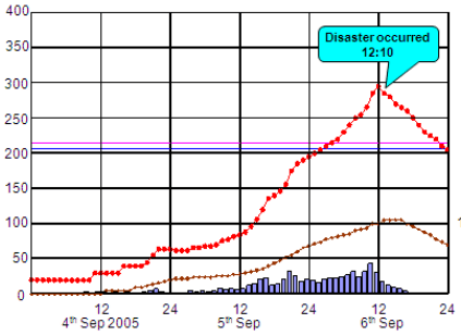


Figure 1.3: Rouge : indice d'eau dans le sol; Brown: précipitations cumulées sur les dernières 24 heures; barres: précipitations sur 1 heure

1.2 Régression linéaire

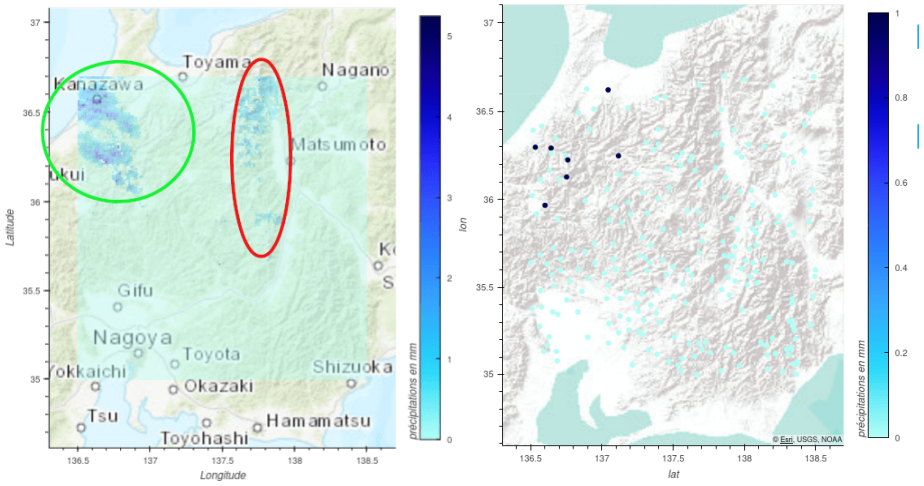
À partir des données radar et jauge, l'objectif est de développer un modèle de machine learning capable d'estimer la pluie. Il ne s'agit pas ici de prévoir la pluie, mais simplement d'établir un modèle capable déterminer les quantités de précipitations à un instant t, à partir de mesures radar passé, présente et future.

1.2.1 Données

Nous disposons de deux types de données collectées au coeur de l'île principale du Japon sur les 3 premiers mois de 2010:

- Radar (fig 1.4a) : Les données radars sont prisent à intervalle temporel de 10 minutes. Les radars quadrillent la surface avec une dimension de $204 * 160 = 21624$. Chaque mesure représentent la quantité de précipitations sur les 10 dernières minutes.
- Jauge (fig 1.4b) : Les données de jauge sont récupérées toute les heures. Un total de 371 jauge couvrent cette surface. Les mesures représentent la quantité moyenne de pluie tombée toute les 10 minutes sur la dernière heure. Pour chaque heure nous avons donc une seule valeur.

Cette zone est intéressante car elle contient les alpes Japonaises, une chaîne de montagne traversant le coeur du Japon. Nous pouvons voir que des précipitations de quelques millimètres ont lieu dans ces montagne (encadré en rouge). Vers Kanazawa (encadré en vert) le vent froid provenant de Siberie est humidifié par la mer du Japon (située du côté de la Chine et la Corée). Ce vent est ensuite bloquée par les montagnes générant des nuages de précipitations.



(a) Radar (b) Jauge

Figure 1.4: Mesure des précipitations le 2 janvier 2010 au Japon

Analyse des données

La première étape en machine learning consiste à analyser ses données. Sur le graphe 1.5 on peut comparer les mesures du radar et de la jauge en un emplacement. les mesures sont toujours exprimées en millimètres mesure sur

une certaine période. 1 millimètres de pluie correspond à 1 litre par mètre carré. On remarque que les mesure ne correspondent pas parfaitement. Alors que la jauge mesure des faible précipitions durant la journée du 13 février, le radar ne semble rien mesurer. La jauge est une mesure plus fiable que le radar à cet emplacement, on peut donc supposer que des interférences on empêchée le radar de mesurée la pluie à cet emplacement.

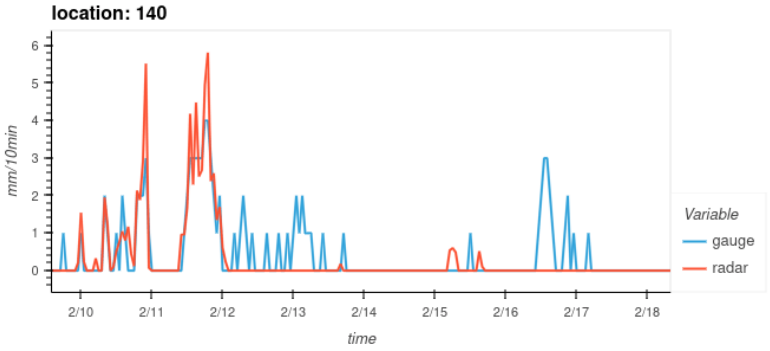


Figure 1.5: Quantité de pluie mesurée à un emplacement contenant une jauge

Pour aller plus loin nous cherchons à observer les mesure de radar potentiellement fausse. Par exemple si le radar mesure une forte précipitations à un emplacement, mais que les 8 emplacements voisins l'entourant ne mesure aucune pluie alors on peut estimer que la forte précipitation mesurée est fausse. Pour trouver ces mesures aberrantes nous calculons à chaque emplacements la plus petite différence de pluie entre cet emplacement et ces 8 voisins l'entourant. Si cette valeur est très élevé (ou très petite) on peut supposer que l'erreur provient du radar. Sur la figure 1.6 par exemple, la mesure de 25mm de pluie ne fait pas de sens.

L'une des caractéristiques de nos données comme nous pouvons le voir ici, est que la distribution des quantités de précipitation par tranche de 10 minutes n'a rien de gaussien. En effet 90 pourcent des données auront pour valeur 0 a 1 millimètres de pluie, et seuls quelques extrêmes se démarqueront avec des précipitations de 20 millimètres. Le traitement de valeurs extrêmes est généralement plus complique dans notre cas. En effet même si 20 millimètres de pluie est une valeur normale, elle peut paraître extrême compare à la quantité énorme de valeurs proche de 0.

Les données de jauge non-plus ne sont pas parfaite, de nombreuses jauge ne semblent pas fonctionner car elles ne mesurent aucune pluie durant 3 mois. Certaine jauge nécessitent une intervention humaine pour récupérer les données de précipitations, d'autre sont automatiquement mesurée par outil d'enregistrement. Les erreur de mesure peuvent provenir de ces deux facteurs.

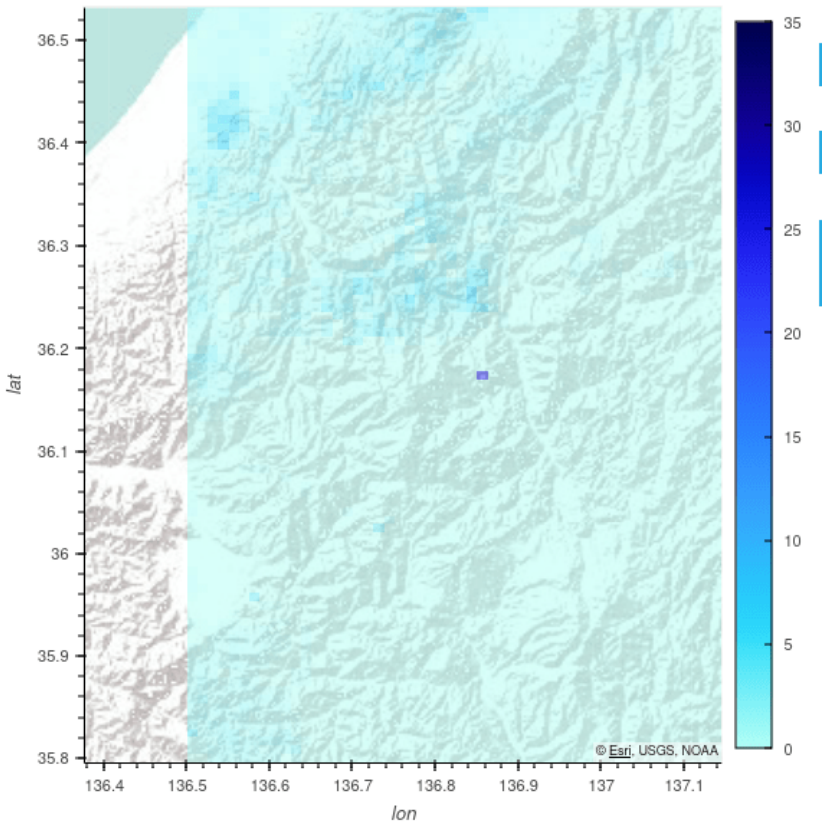


Figure 1.6: Valeurs aberrantes

Traitement des données

Les données de jauge n'ayant aucune valeur sont simplement évincées, ces jauges ne seront donc pas utiliser dans notre modèle de machine learning. Les valeurs aberrantes sont réduites aux valeurs de leur plus proche voisin. On considère une valeur aberrante lorsque la mesure effectuée en un point est 20 mm plus élevé que ses 8 plus proches voisins.

Pour définir un modèle de machine learning il faut définir les données entrant dans ce modèle et celle sortant. Dans tout les modèles que nous testerons, les données entrantes seront les valeurs mesurée par le radar. Ces données sont normaliser afin d'être toute compris entre 0 et 1. Cela permet de mettre toute les donnée sur une même échelle. Les données rentrant seront du même format que les données entrante.
Il faut ensuite définir les exemples annotés servant de base d'apprentissage pour notre modèle. Les exemples annotés seront les valeurs mesurée par les jauges toute les heures.

Nous allons créer différent modèle de régression linéaire ayant pour objectif d'estimer la pluie tombée au niveau des jauges a partir des données radars. Nous ne prendrons les données que d'une jauge au départ afin de faciliter le problème, puis les modèles linéaires suivant utiliserons les données de toute les jauges. Les données en entrées sont donc les mesures radars effectue au niveau de la jauge toute les 10 minutes. L'objectif est d'obtenir une estimation heure par heure proche des valeurs de jauge.

1.2.2 Regression lineaire sur une jauge

Avant même de créer notre premier modèle de régression linéaire, mous allons construire un modèle simple qui ne demande aucune connaissance et machine learning. Il s'agit simplement de calculer la moyenne des données radar pour chaque heure. Il nous serviras de base de comparaison pour évaluer les modèles linéaires.

Définition du modèle linéaire

Nous créons ensuite un modèle de régression linéaire f.

$$f(X) = a * x0 + b * x1 + c * x2 + d * x3 + e * x4 + f * x5 \tag{1.1}$$
$$f(X) = \Theta X \tag{1.2}$$

Où :

xi = valeur du radar a la minute 0, 10, 20... aussi appelé features
a, b, c, d, e, f = paramètres de notre modèle
$X^T = \begin{pmatrix} x0 & x1 & x2 & x3 & x4 & x5 \end{pmatrix}$
$\Theta = \begin{pmatrix} a & b & c & d & e & f \end{pmatrix}$

Puis une fonction coût J permettant d'évaluer les prédictions de notre modèle linéaire. La fonction que nous utilisons est l'erreur quadratique moyenne calculée par :

$$J(\Theta) = \frac{1}{n} \sum_{i=1}^{n} \left(y_{i\_jauge} - f(X_i) \right)^2$$

Où :

n = le nombre d'échantillons dans notre jeu d'entraînement.
$f(X_i)$ = correspond à la valeur obtenu par notre modèle sur le i_ème échantillon.
y_{i_jauge} = la valeur mesurée par la jauge, c'est le label associé a X_i.

La fonction coût ne dépend que des paramètres (Θ) de notre modèle. Plus cette fonction est petite pour un échantillon X_i, plus les prédictions de notre modèle se rapproche des valeurs idéales de jauge. Nous cherchons donc à obtenir la meilleur combinaison de paramètre Θ tel que la fonction coût soit le plus petit possible sur le jeu d'entraînement de notre modèle.

La méthode que nous utilisons pour trouver les meilleurs paramètres Θ est un algorithme d'apprentissage supervisé nommée descente du gradient [2]. On appelle apprentissage supervisé un algorithme apprenant à associe un input x à un output y, en s'entraînant sur un jeu de donné de composé d'input labélisé. Le modèle est entraîné sur ces input labélisé durant un certain nombre l'itération. À chaque itération du modèle sur notre jeu d'entraînement nous calculons $f(X)$ et nous comparons notre résultat avec le label y_{i_jauge} en calculant $J(\Theta)$. À chaque itération les paramètres Θ du modèle sont légèrement modifié afin de minimiser la fonction coût lors de la prochaine itération. Cette étape est réitéré pendant un certain nombre d'époque, jusqu'à que la fonction coût est atteint son minimum.

Évaluation du modèle sur la jauge d'entraînement

Le modèle de régression linaire est entraîné sur le jeu de donnée d'entraînement pendant 100_000 époque. Dans la figure 1.7 on peut observer l'évolution décroissante de la fonction coût, jusqu'à atteindre une limite. Lors de l'entraînement il est important de calculer la valeur de la fonction coût sur les données d'entraînements et sur les données test. Cela permet de vérifier que notre modèle n'est pas en sur-apprentissage. En effet sur la fonction coût des données de test commence à croître alors que celle des données d'entraînement suit une allure régulière, cela signifie que le modèle s'est excessivement entraîner à représenter le jeu d'entraînement et n'est plus capable de prédire d'autre données.

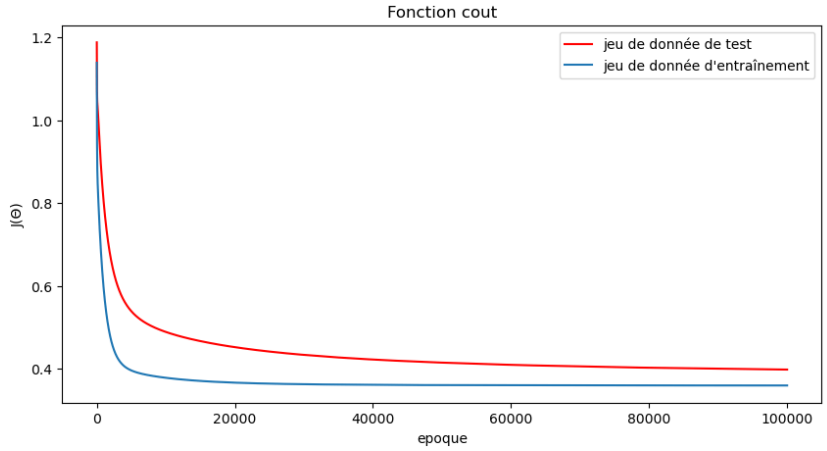


Figure 1.7: Évolution de la fonction coût

La figure 1.8 nous montre les résultats obtenus par notre modèle (trait plein bleu) en comparaison du modèle simpliste calculant la moyenne (tiret orange). La courbe rouge est la valeur idéale de jauge. On remarque que notre modèle linéaire n'est pas plus efficace que la valeur moyenne et semble faire les mêmes erreurs de prédiction. En appliquant notre modèle sur toute les jauges et en

comparant sa précision à la moyenne des radars , on se rend compte que le modèle est 15% moins précis que la moyenne des radars. Sur la figure 1.8 on peut observer certain défaut de notre modèle. Le modèle de régression linéaire n'arrive pas a représenter la valeur minimum de 0 et semble avoir une valeur constante aux alentours de 0.1mm de pluie. De plus, les fort pics de pluies sont sous-évaluer par la régression linéaire. Ces problèmes proviennent du fait que notre modèle est une fonction linéaire tentant de représenter un problème de plus haut degré polynomiale.

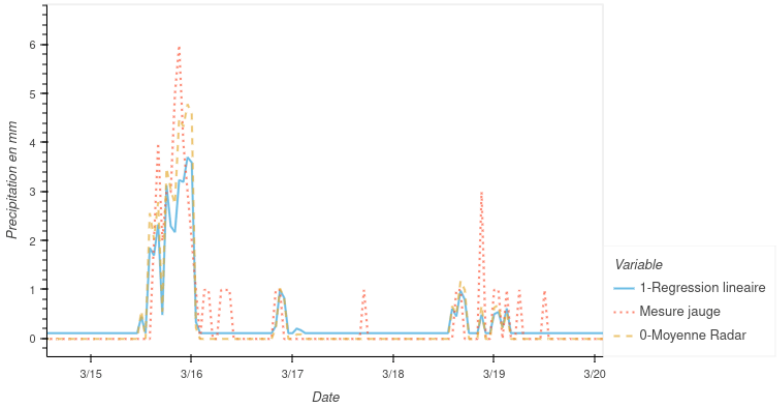
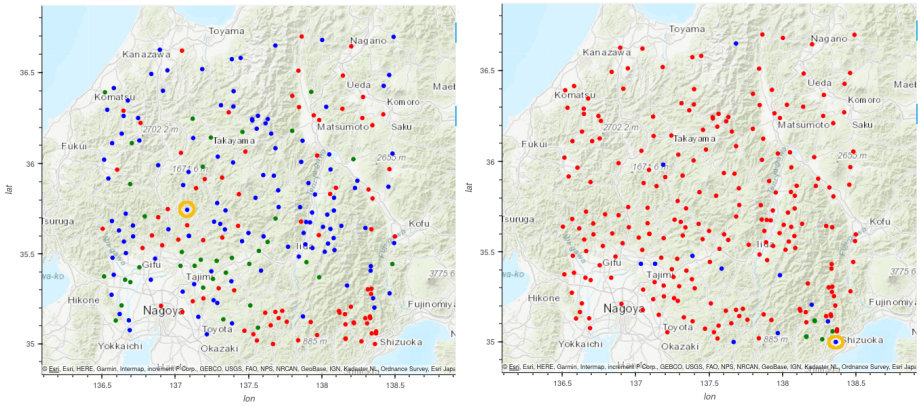


Figure 1.8: Mesure moyenne des radars comparé à la régression linéaire

Évaluation du modèle sur toute les jauges

Le modèle à été entraînée sur une jauge. Le but du modèle est d'être généralise à n'importe quelle mesure de jauge et de radar. Nous évaluons donc les performances de notre modèle sur des jauge qui lui sont totalement inconnus. Si les données d'entraînements correspondent à un zone géographique plane, lorsque des données provenant d'une zone montagneuse sont inséré il y a de forte chance que les résultats soient mauvais. La figure ci dessous compare le modèle lorsqu'il est entraînée dans une zone montagneuse au bord de la mer 1.9b, et un entraînement sur une zone retranchée dans les terres 1.9a. Les points en vert représentent les jauges ou le modèle fait une meilleur prédiction que une simple moyenne. Les points en bleu et rouge représentent respectivement une prédiction à peu près égale à la moyenne, ou une meilleur prédiction de la part de la moyenne. On peut voir que en fonction de la position des données d'entraînements les prédictions changent drastiquement. Lorsque le modèle est entraîné dans les terres, il n'arrive pas à prédire les précipitations dans le Sud-Est, zone montagneuse au bord de la cote. Au contraire lorsque le modèle est entraînée dans le Sud-Est, quelques bonne prédictions sont réalisées dans cette zone montagneuse. Cependant le reste des prédictions sont fausse. On dit que les covariables sont décalées : la distribution du jeu de donnée d'entraînement

ne correspond pas au reste des données de test. La zone au Sud-Est a tendance a avoir un temps calme avec de fort pic de pluie, ce qui se démarque des autres climats. Notre modèle ne performe donc pas bien car la distribution des données d'entraînement et de test sont différentes. Pour que notre programme se généralise, nous allons entraîner le modèle sur toute les données de jauge.



(a) Jauge d'entraînement 358 encadrée en orange

(b) Jauge d'entraînement 171 encadrée en orange

Figure 1.9: Performance du modèle. rouge : la moyenne performe mieux que le modèle. Vert : le modèle performe mieux. Bleu : moyenne et modèle ont a peu près les mêmes performances.

1.2.3 Entraînement sur toute les jauges

Jusqu'à maintenant le modèle de régression linéaire ne s'entraînait que sur les données d'une jauges. Nous allons maintenant l'entraîner sur tout les emplacements de jauges, ce qui permet au modèle de s'exercer sur des données plus variées et représentative de la réalité. En entraînent notre modèle de cette manière sa précision dépasse celle de la valeur moyenne de peu. La figure 1.10 nous montre les performances du modèle par rapport à la moyenne. Contrairement aux résultats de la section précédente le modèle arrive à égaler ou dépasser la moyenne sur la plupart des points d'évaluations. On remarque cependant que la zone située au Sud-Est n'obtient toujours pas de bon résultats avec le modèle. En comparant les mesures dans cette zone avec les autres jauge, on voit que les pics de pluie sont plus puissant. En regardant la quantité de pluie tombée sur les 3 mois d'observation, on observe un surplus de 20% dans la zone sud-est.

En analysant la fonction linéaire que l'on a entraînée, on remarque aussi que les coefficients Θ semblent tendre vers des valeurs proche de 1/6 1.11. Le modèle linaire calcul donc la moyenne des valeurs radars pour chaque heures,

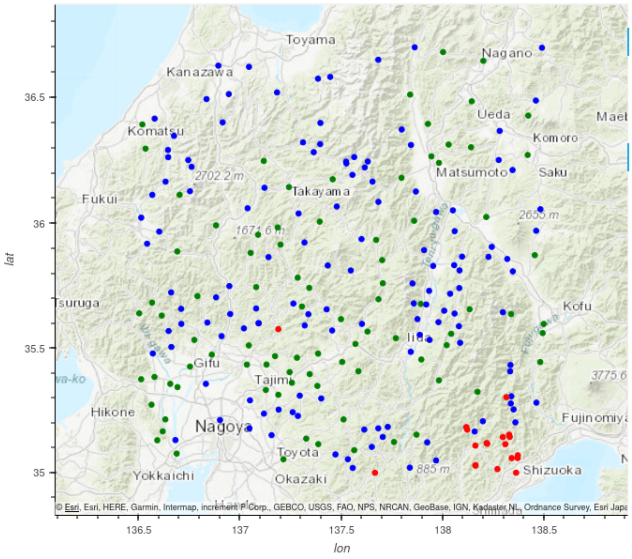


Figure 1.10: Performance du modèle entraîné sur toutes les jauges. rouge : la moyenne performe mieux que le modèle. Vert : le modèle performe mieux. Bleu : moyenne et modèle ont a peu près les mêmes performances.

cela confirme que la fonction moyenne est la fonction linaire la plus efficace pour ce problème. En réalité les valeurs Θ ne tendent pas exactement vers 1/6, mais vers une valeur légèrement plus faible. De plus le modèle semble donner plus d'importance au coefficient représentant la valeurs du radar a 50 minutes. Notre modèle est donc une valeur moyenne atténué donnant plus d'importance à un coefficient.

La régression linéaire est un premier pas vers des modèles plus compliquées du machine learning. Grâce à ce modèle nous pouvons avoir une première estimations des précipitations sur le territoire japonais. Un tel modèle est très simple mais nous a permis de démontrer l'importance de certaine caractéristique permettant de prédire les précipitations, comme la prise en compte des dimensions temporels lorsque l'on estime une précipitation. Cependant ce modèle n'est pas assez complexe pour estimer une système non-linéaire.

1.3 Importance des caractéristiques temporelles pour estimer les précipitations

Les données utilisées pour le modèle linéaire sont les mesures radars sur l'heure que l'on estime. Ces mesures sont nommées caractéristiques ("features" en anglais). Pour améliorer notre modèle l'ajout de caractéristique est nécessaire.

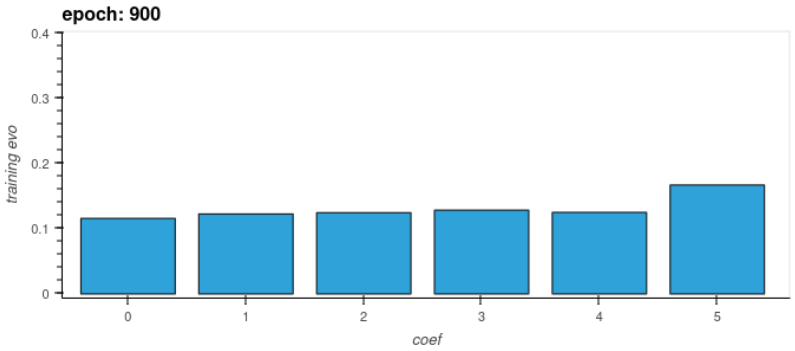


Figure 1.11: Coefficients finaux obtenus par la régression linéaire. Les coefficients représentent les mesures du radar toute les 10 minutes durant l'heure estimée

Par exemple l'ajout d'informations spatiales ou de données atmosphériques permettent d'améliorer les performance d'un modèle. Dans cette section nous étudions l'importance de caractéristique temporels.

1.3.1 Modèle linéaire avec une fenêtre temporelle

Pour estimer la pluie à une certaine heure, les modèles précédents utilisent les 6 données radars mesurées sur cette heure. Dans le modèle suivant l'objectif est d'étendre les données temporel que nous utilisons. Pour une estimation a une certaine heure H nous utilisons donc les données radars liées à l'heure H, mais aussi H+1 et H-1. Dans ce cas nous avons une fenêtre temporelle de taille 3.

Une fois de plus, la précision de notre modèle s'améliore légèrement. Sur la figure 1.12 on peut observer l'évolution de la fonction coût en fonction de la fenêtre temporelle choisis. Il est intéressant de remarquer que lorsque la fenêtre temporelle devient trop large, les performances du modèle diminue. Le surplus de données empêche les plus importantes de se démarquer. En effet toute les caractéristiques temporelles ne se valent pas. Par exemple les valeurs de pluie ayant eu lieu en H+100 minutes sont beaucoup moins importante que celle ayant lieu en H. La figure 1.13 montre les coefficients obtenus par la régression linéaire. Les coefficients sont ordonnées de -7H à +7H par tranche de 10 minutes. Par exemple, le coefficient -7 représente 70 minute avant la mesure de la jauge en H. En s'entraînant le modèle a de lui même atténuer les coefficients représentant des heures éloignés de l'heure H. On voit sur cette figure que les coefficients les plus important sont situés entre -10 et 10 (soit H-100 minutes et H+100 minutes), ce qui permet au mesures temporel adjacentes à notre heure H d'impacter fortement le résultat de la régression.

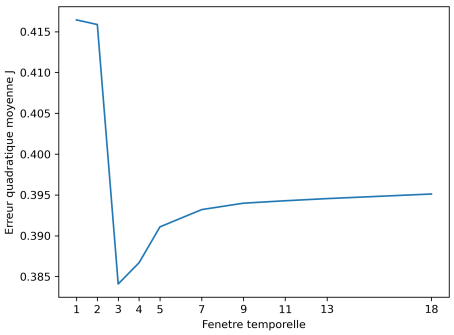


Figure 1.12: Évolution de la fonction coût en fonction de la taille de la fenêtre temporelle

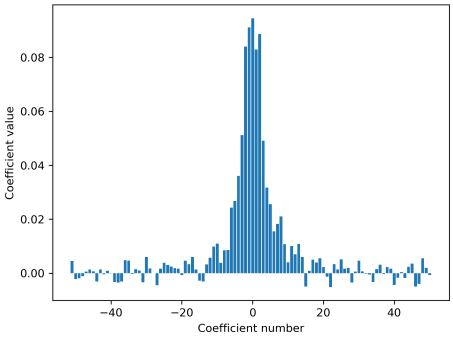


Figure 1.13: Coefficient du modèle pour une fenêtre de taille -7H, +7H

1.3.2 Analyse des caractéristique avant l'entraînement

Précédemment nous avons analyser l'importance de nos caractéristiques après avoir entraînées notre modèle. Dans le cas où l'entraînement de notre modèle est long, obtenir des informations sur nos caractéristiques en amont de l'entraînement est nécessaire. Pour cela il existe différentes méthode d'analyse des caractéristiques tel que la corrélation par rapport a la cible, et la PCA (Principal Component Analysis).

Corrélation avec la cible

Le coefficient de corrélation est une mesure quantifiant la relation linéaire entre 2 variables. Dans notre cas ce coefficient est utiliser pour calculer la corrélation entre une caractéristique et la cible de notre modèle. Pour chaque mesure du radar au temps H, H+10 minutes, H-10 minutes etc.. nous nous observons la corrélation avec les mesures de la jauge au temps H. La figure 1.14 montre le coefficients de corrélations obtenus. Un coefficient de corrélation est compris entre -1 et 1. Plus la valeur est proche de 1 (ou -1), plus les variables sont positivement (ou négativement) corrélées, c'est à dire dire qu'elles évoluent de la même manière. On peut voir que les caractéristiques temporelles au temps H à H+6 ont les coefficients de corrélations les plus élevées. Plus l'ont s'éloigne des caractéristiques au temps H, moins les variables sont linéairement corrélées. Pour réduire le nombre de caractéristiques que l'on choisis il suffit de définir un seuil de corrélation. Toute caractéristique ayant un coefficient plus faible que le seuil sont rejetées.

PCA

Principal Component Analysis est une méthode permettant de condenser l'ensemble des caractéristiques envoyer en input du modèle. Dans notre cas l'objectif est de

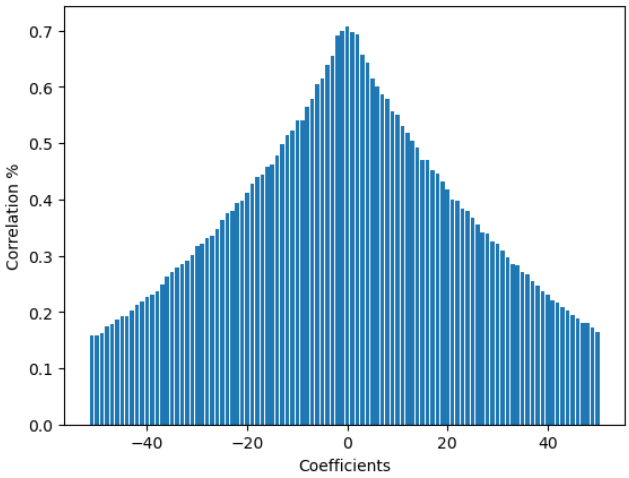


Figure 1.14: Coefficients de corrélations entre les caractéristiques et la cible du modèle

réduire les caractéristiques choisis en un ensemble de 6 nouvelles caractéristiques qui seront une combinaison linéaire des précédente. On peut voir la PCA comme une réorganisation de nos caractéristiques afin de les ordonnée en fonction de la quantité d'information qu'elle fournissent. Contrairement aa l'exemple precedent ou l'

- caracteristiques temporel des precipitation (coef, pca, correlation)

Lors de L'etude pca : consisterait a selection les caracteristiaue les plus importantes de notre jeu de donnees, et utiliser ces caracteristiques en input de la regression

- limitation modele lineaire Ajout de plus de caractéristique (les donnees voisines, un plus large temporel ...) impossible pour le modele lineaire

- Limitation dans les features choisis pour le modele On n'utilise pas les informations spatiale autour de notre point. C est pour ca qu'on voit que notre modele lineaire ne fait que copier les donnees radars : la ou le radar mesure bcp de pluie, le modele lineaire aussi. (Mettre une image ou le radar et la pred monte, alors que la jauge ne mesure pas de pluie)

Bibliography

[1] *Japan Meteorological Agency*. URL: https://www.jma.go.jp/jma/en/copyright.html.

[2] Machine Learnia. *DESCENTE DE GRADIENT (GRADIENT DESCENT) - ML#4*. 2019. URL: https://www.youtube.com/watch?v=rcl_YRyoLIY.

[3] Kazuhiko Nagata. "Quantitative precipitation estimation and quantitative precipitation forecasting by the Japan Meteorological Agency". In: *RSMC Tokyo–Typhoon Center Technical Review* 13 (2011), pp. 37–50.

[4] Ryota Nakamura and Yukihiro Shimatani. "Extreme-flood control operation of dams in Japan". In: *Journal of Hydrology: Regional Studies* 35 (2021), p. 100821. ISSN: 2214-5818. DOI: https://doi.org/10.1016/j.ejrh.2021.100821. URL: https://www.sciencedirect.com/science/article/pii/S2214581821000501.

[5] Satoru Oishi, Toshihiko Tahara, and Mariko Ogawa. "Study on optimization of the operation of dams using ensemble prediction and a distributed rainfall-runoff model". In: *HIC*. 2018, pp. 1584–1588.