

TD Solar Energy

Description:

Dans ce laboratoire, vous utiliserez vos nouvelles connaissances pour analyser un jeu de données réel lié à l'énergie solaire (voir Figure 1). Pour réussir ce TD, vous devrez importer les données et répondre à des questions en utilisant des outils de manipulation et d'analyse de données.

La Figure 1 présente une vue d'ensemble du processus de production d'électricité, depuis un module de panneaux solaires jusqu'au réseau électrique. L'énergie solaire est directement convertie en électricité grâce à l'effet photoélectrique. Lorsque des matériaux tels que le silicium (le matériau semi-conducteur le plus couramment utilisé dans les panneaux solaires) sont exposés à la lumière, des photons (particules subatomiques d'énergie électromagnétique) sont absorbés. Cela provoque la libération d'électrons libres et, par conséquent, la production d'un courant continu (DC).

En pratique, à l'aide d'un onduleur, le courant continu est converti en courant alternatif (AC), puis injecté dans le réseau électrique, où il peut être distribué aux habitations. Il existe différents types de systèmes solaires, tels que les systèmes hors réseau (off-grid) et les systèmes photovoltaïques directs, illustrés dans la Figure 2.

Le jeu de données utilisé dans ce TD est un jeu de données personnalisé, construit à partir de mesures de rayonnement solaire provenant du Measurement and Instrumentation Data Center (MIDC) du National Renewable Energy Laboratory (NREL) des États-Unis.

La station sélectionnée est située à l'Université du Nevada à Las Vegas (UNLV), et les données exploitées couvrent une période d'un an, entre 2021 et 2022. De plus, ce jeu de données possède une résolution d'une minute et contient 21 variables liées aux conditions météorologiques et à d'autres paramètres pertinents, notamment :

- la température ambiante,
- la vitesse du vent,
- la direction du vent,
- l'irradiance globale horizontale (GHI),
- l'irradiance normale directe (DNI),
- l'irradiance diffuse horizontale (DHI),
- les angles zénithal et azimuthal,
- la masse d'air,

- ainsi que d'autres variables.

L'objectif principal de ce travail est d'analyser le niveau de rayonnement solaire en fonction de la date et de l'heure, à partir des mesures de l'année précédente.

Cette étude peut être utile pour déterminer si l'utilisation de batteries de stockage d'énergie solaire (stockage et distribution de l'énergie, comme illustré dans la Figure 3) sera pertinente à l'avenir. Cela est possible en prédisant le niveau de rayonnement solaire.

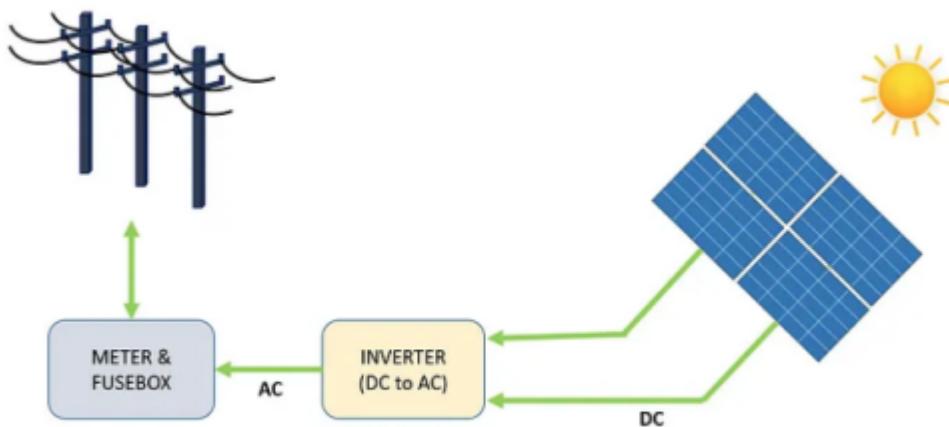


Figure 1: Grid connected Solar System

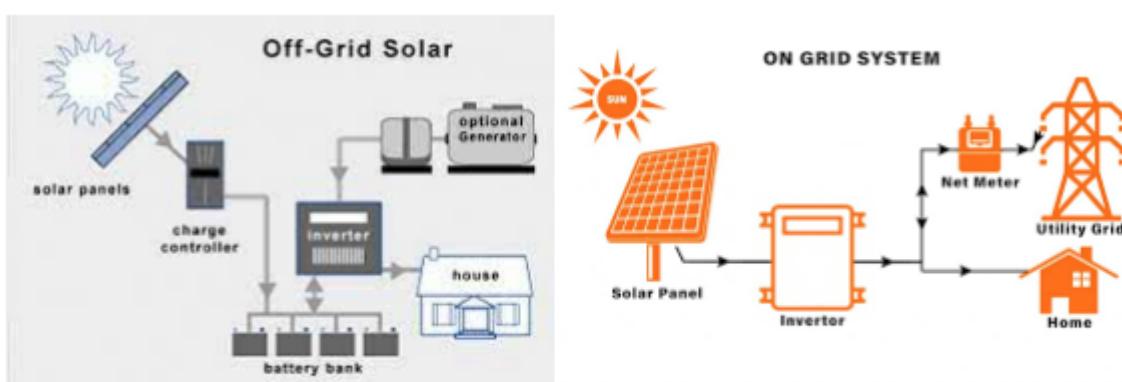


Figure 2: Types of solar power systems



Figure 3: An example of an implemented solar energy system

Questions :

1.1 Chargement des données

- Votre première tâche consiste à importer les bibliothèques Python nécessaires pour réaliser ce TD, telles que Pandas et Numpy.
- Ensuite, chargez le fichier de données correspondant ([solar2021UNLV.csv](#)) dans un DataFrame nommé df_solar. Quelle colonne devez-vous utiliser comme index ? Pourquoi ?

1.2 Exploration des données

Dans cette partie, nous allons examiner le jeu de données afin de vous permettre de mieux comprendre les informations dont vous disposez.

1. Quelle est la taille du DataFrame chargé ?
2. Affichez les 10 premières lignes et les 10 dernières lignes du DataFrame df_solar.
3. Listez les noms des colonnes présentes dans ce jeu de données ainsi que leur type de données.
4. Quelles variables (features) peuvent contenir des valeurs aberrantes (outliers) ? Vous devez justifier votre réponse à l'aide d'un graphique, par exemple un boxplot.

1.3 Manipulation des données

5. Comme vous avez pu le remarquer, le DataFrame df contient une ou plusieurs colonnes inutiles. Supprimez la ou les colonnes non souhaitées.

- 6.** Vérifiez la présence de valeurs manquantes (NaN) dans le DataFrame df_ref, puis supprimez toutes les lignes contenant des valeurs manquantes.
- 7.** Vérifiez la présence de doublons dans ce DataFrame, puis supprimez toutes les lignes dupliquées.

1.4 Analyse des données et visualisation

- 8.** Quelle est la valeur maximale de la variable “GHI” et à quelle date correspond-elle ?
- 9.** Quelle est la valeur moyenne de “GHI” :
 - globale ?
 - par mois ?
 - par saison ?
- 10.** Tracez l’évolution de “GHI” pour une journée donnée.
Puis, tracez un second graphique pour cinq jours consécutifs à partir d’une date choisie.

Vous pouvez extraire les valeurs de GHI pour une période donnée à partir de la série temporelle en utilisant le slicing suivant :

```
df_ref['2021-06-01':'2021-06-02']['ghi']
```

- 11.** Tracez la variation de “GHI” par jour et par mois.
Pouvez-vous réaliser cela sans ajouter de nouvelles colonnes correspondant au jour et au mois ? Si oui, expliquez comment.
- 12.** Tracez l’évolution de :
 - GHI**
 - DHI**
 - DNI**
 - la température**

en fonction de la date (jour), sous forme de sous-graphiques (subplots), un graphique par variable.

- 13.** Tracez la variation de la température en fonction du rayonnement solaire “GHI”.
Que pouvez-vous conclure sur la relation entre ces deux variables ?

14. Tracez la matrice de corrélation du jeu de données sous forme de carte thermique (heatmap) afin de mieux comprendre les relations entre toutes les variables du dataset.

- Utilisez la fonction `corr()` de Pandas pour calculer la matrice de corrélation.
- Stockez le résultat dans une variable appelée `corr`.

Ces valeurs permettent d'analyser les relations entre les colonnes :

- Une corrélation parfaitement positive est égale à +1.
- Une corrélation parfaitement négative est égale à -1.

Ensuite, répondez aux questions suivantes :

- En observant la diagonale principale (de gauche à droite) de la matrice de corrélation, pourquoi cette diagonale est-elle remplie de valeurs égales à 1 ? Expliquez.
- Toujours en observant la matrice de corrélation, vous remarquerez que les valeurs sont symétriques : les valeurs situées sous la diagonale ont un équivalent au-dessus. Pourquoi la matrice de corrélation est-elle symétrique ? Expliquez.
- De nombreuses paires de variables présentent une corrélation proche de zéro. Que signifie une corrélation proche de zéro ?
- Quelles sont les variables qui présentent les corrélations les plus fortes ? Pouvez-vous les lister ?

15. Tracez un nouveau graphique de corrélation entre le rayonnement “GHI” et les autres variables du dataset.

- Rappelez que la fonction `corr()` utilise par défaut la méthode de corrélation de Pearson.

Répondez ensuite aux questions suivantes :

- Quelles variables présentent la plus forte corrélation avec “GHI” ? Listez-les.
- Quelles variables ne présentent pas ou peu de corrélation avec “GHI” ? Listez-les.
- Reproduisez la même analyse en utilisant d'autres méthodes de corrélation disponibles dans Pandas :

- le coefficient de corrélation de Kendall (Kendall Tau),
- le coefficient de corrélation de Spearman.
- Quelles conclusions pouvez-vous tirer en comparant les résultats obtenus avec les méthodes Pearson, Kendall et Spearman ?

16. Créez un nuage de points (scatter plot) pour illustrer la relation entre :

- l'angle d'azimut (*Azimuth Angle*) en abscisse,
- le rayonnement GHI en ordonnée.

Observez-vous un motif ou une structure particulière dans ce graphique ? Justifiez votre réponse.

17. Créez un nuage de points pour analyser la relation entre :

- l'angle zénithal (*Zenith Angle [degrees]*) en abscisse,
- le rayonnement GHI en ordonnée.

Existe-t-il une tendance ou un comportement remarquable dans ce graphique ? Expliquez.

18. Créez un nuage de points représentant :

- l'angle d'azimut en abscisse,
- l'angle zénithal en ordonnée,

et utilisez la variable “GHI” comme code couleur pour les points.

- Le résultat obtenu confirme-t-il les conclusions précédentes concernant la relation entre GHI, azimut et zénith ?
- Justifiez votre réponse à l'aide du graphique.

19. Affichez un graphique de densité de probabilité bidimensionnelle (KDE 2D) pour analyser :

- la relation entre le rayonnement GHI et la température.

Ensuite, réalisez un autre KDE 2D pour étudier la relation entre :

- la vitesse du vent (*wind speed*),
- la direction du vent (*wind direction*).

20. Tracez la distribution des variables suivantes sous forme de sous-graphiques (subplots) :

- la vitesse du vent,
- la direction du vent,
- la température.

21. Tracez un graphique KDE bidimensionnel (2D) illustrant la distribution conjointe des variables :

- DHI (*Diffuse Horizontal Irradiance*),
- DNI (*Direct Normal Irradiance*).

1.5 Investigation des différences

22. Créez une figure unique contenant les courbes KDE (qui montrent la densité) de ‘DHI’ et ‘DNI’.

23. Fournissez un pair plot qui représente toutes les relations par paires dans le dataset en utilisant la méthode correspondante de la bibliothèque Seaborn.
Que pouvez-vous en conclure ?

1.6 Investigation supplémentaire

Créez un nuage de points 3D avec les variables (x, y, z) : où ‘DNI’, ‘DHI’ et ‘GHI’ représentent respectivement les axes x, y et z.

Vous pouvez utiliser le code suivant :

```
ax = plt.axes(projection='3d')  
  
ax.scatter3D(x, y, z)
```

- Que pouvez-vous conclure à partir de ce graphique ?

1.7 Échantillonnage descendant et ascendant des séries temporelles

Lors de l'évaluation des ressources solaires, il se peut que vous ayez besoin d'une résolution temporelle différente de celle de vos données pour une partie particulière de l'analyse. Dans ces cas, il est possible de rééchantillonner à la baisse (down-sampling) ou à la hausse (up-sampling) les données à différentes résolutions temporelles en utilisant deux méthodes de la bibliothèque Pandas appelées `resample` et `asfreq`.

Selon vos besoins, vous choisirez l'une ou l'autre méthode.

Quelle que soit la méthode utilisée, les deux nécessitent un DataFrame avec un DatetimeIndex, soit conscient du temps (time-aware / localisé), soit non conscient du temps (time-naive / non localisé).

Quelle est la différence entre ces deux types ? Essayez d'expliquer la différence en détail.

- Tracez la variation de 'GHI' pour des intervalles de 30 minutes, 1 heure et 1 semaine. Que pouvez-vous en conclure ?
- Tracez la variation de 'GHI', 'DHI', 'DNI' et de la température pour des intervalles de deux semaines sous forme de subplots.