

Projet de mathématiques et signal

Identifier l'existence de cycles dans des données météorologiques

I. Introduction	2
II. Description statistiques des données	2
A. Statistiques descriptives des variables météorologiques	2
B. Dispersion des données	3
1. <i>Température moyenne</i>	3
2. <i>Précipitations</i>	3
3. <i>Pression</i>	4
4. <i>Durée d'ensoleillement</i>	4
III. Outils mathématiques utilisés	5
IV. Cycles annuels, synoptiques et pluri-annuels	6
A. Température moyenne, minimale et maximale	6
B. Précipitations	7
C. Chutes de neige	8
D. Vitesse du vent et pics de rafales	9
E. Pression	10
F. Ensoleillement	11
V. Conclusion	11

I. Introduction

Dans le cadre de ce projet, nous disposons des relevés journaliers de la station météorologique de Strasbourg-Entzheim couvrant la période 1950-2024, et nous allons étudier les variations des signaux de données au cours du temps. Ces données incluent les températures (moyenne, minimale et maximale), les précipitations, l'épaisseur de neige, la vitesse et la direction du vent, la pression atmosphérique ainsi que la durée d'ensoleillement.

Cette étude se fera en deux étapes, d'une part une description statistique des données pour caractériser leur distribution et identifier les valeurs extrêmes, et d'autre part, utiliser des outils mathématiques et d'analyse fréquentielle afin de détecter les cycles annuels, synoptiques et pluri-annuels. Le tout sera basé sur des modélisations fréquentielles et temporelles adaptées aux cycles que nous cherchons à identifier. Nous allons aussi évaluer la proportion de bruit dans les séries.

II. Description statistiques des données

Préalablement à l'analyse temporelle et fréquentielle, il est nécessaire de réaliser une description statistique des données, afin de caractériser globalement les relevés météorologiques étudiés. Cette étape vise à analyser la distribution des variables, ainsi que la présence de valeurs extrêmes. La variable wdir n'a pas pu être exploitée en raison de l'absence de données.

A. Statistiques descriptives des variables météorologiques

Voici tout d'abord une analyse statistique descriptive des différentes variables à l'aide de la fonction `describe()` de la bibliothèque Python *Pandas*.

	time	tavg	tmin	tmax	tsun
count	27394	27394.000000	27394.000000	27394.000000	14429.000000
mean	1987-07-02 12:00:00	10.691911	6.331576	15.273980	297.074919
min	1950-01-01 00:00:00	-14.800000	-23.200000	-14.400000	0.000000
25%	1968-10-01 06:00:00	5.100000	1.300000	8.500000	30.000000
50%	1987-07-02 12:00:00	10.900000	6.600000	15.500000	234.000000
75%	2006-04-01 18:00:00	16.600000	11.700000	22.200000	522.000000
max	2024-12-31 00:00:00	29.600000	23.800000	38.900000	930.000000
std	NaN	7.459684	6.666170	8.879623	273.329628
	prcp	snow	wdir	wspd	wpgt
count	27394.000000	23388.000000	0.0	18498.000000	16093.000000
mean	1.713069	2.042329	NaN	10.129300	30.403958
min	0.000000	0.000000	NaN	0.000000	1.800000
25%	0.000000	0.000000	NaN	6.000000	20.500000
50%	0.000000	0.000000	NaN	8.900000	28.100000
75%	1.600000	0.000000	NaN	13.000000	37.800000
max	66.300000	400.000000	NaN	93.600000	143.600000
std	3.949863	14.185158	NaN	5.636587	14.029083
	pres				

Avec :

- **Count** : nombre de valeurs non nulles sur la période 1950-2024
- **Mean** : moyenne de la série
- **25% et 75%** : premier et troisième quartiles pour évaluer la dispersion autour de la médiane
- **Std** : mesure de la variabilité ou du bruit dans la série
- **Min/Max** : valeurs extrêmes, pour repérer les épisodes exceptionnels (froids, chaleurs, fortes précipitations)

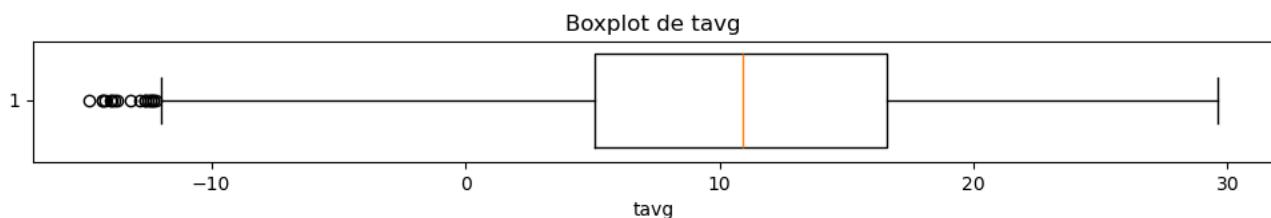
Nous remarquons que certaines variables comme la pression atmosphérique ou l'ensoleillement, présentent de nombreuses données manquantes (on a peu près 27 000 jours entre 1950 et 2024), ce qui peut influer sur la fiabilité des analyses et sur la précision des résultats obtenus. Ensuite, certaines variables présentent des écarts-types très élevés, comme pour le temps d'ensoleillement ou la température moyenne. Cela signifie que ces variables ont une forte variabilité : les valeurs journalières peuvent s'éloigner considérablement de la moyenne, traduisant des fluctuations saisonnières importantes ou des événements climatiques extrêmes.

Par exemple, pour la température moyenne, un écart-type élevé reflète les différences marquées entre hiver et été. De même, pour l'ensoleillement, cela reflète les jours très ensoleillés et les jours nuageux. Ces caractéristiques montrent que, même si la moyenne donne une valeur centrale, elle ne décrit pas entièrement la dispersion et les extrêmes présents dans les données.

B. Dispersion des données

Nous allons analyser des diagrammes moustache sur l'ensemble de la période 1950-2024, afin de visualiser la répartition des données, en mettant en évidence la médiane, les quartiles (25 % et 75 %), les valeurs extrêmes et l'étendue des données.

1. Température moyenne



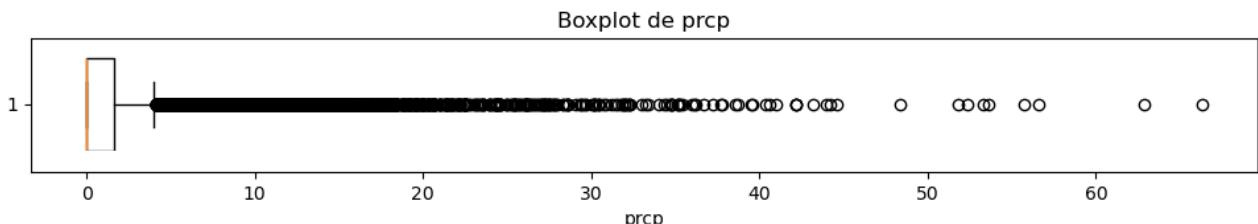
Sur ce diagramme à moustache nous observons que la médiane se situe autour de 10 °C, ce qui correspond à la valeur centrale de l'ensemble des relevés.

Le premier quartile (Q1) est approximativement à 5 °C, tandis que le troisième quartile (Q3) se situe autour de 16 °C, indiquant que 50 % des valeurs journalières sont comprises entre ces deux bornes.

Le minimum et le maximum représentés par les extrémités des moustaches sont respectivement de -15 °C et 30 °C, donnant une idée de l'étendue des températures observées.

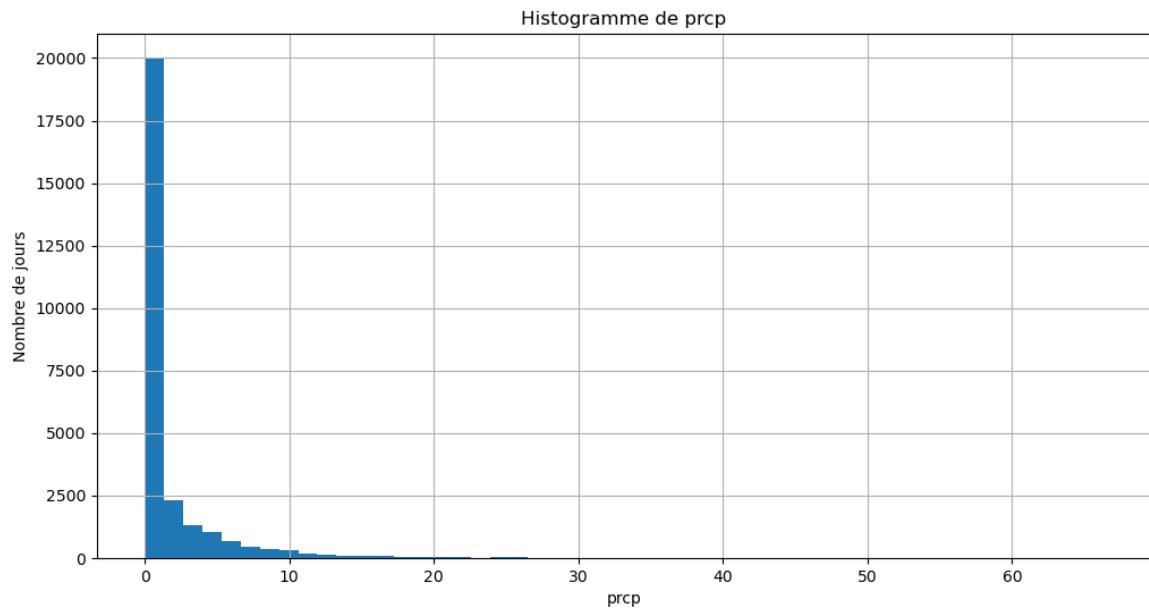
On remarque également la présence de valeurs atypiques en dehors des moustaches, ce qui correspond à des épisodes rares de froid.

2. Précipitations



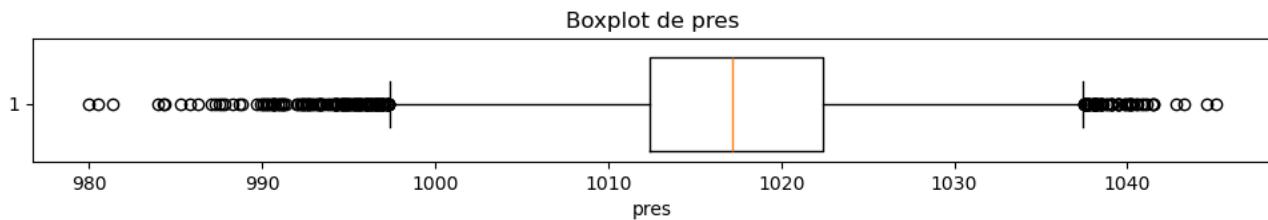
Sur ce diagramme à moustache des précipitations (prcp), nous observons que la médiane, le premier quartile (Q1) et le minimum se confondent et se situent autour de 0 mm, ce qui reflète que la majorité des jours sont sans pluie. Le troisième quartile (Q3) se situe autour de 2 mm, indiquant que plus de 75 % des jours enregistrent des précipitations faibles (0 à 2 mm).

Cependant, nous remarquons la présence d'une grande quantité de valeurs atypiques au-dessus du max. Ces points correspondent à des épisodes de pluie plus intenses, rares mais réels. Cela montre que dans la majorité du temps il pleut peu ou pas, et certains jours de fortes pluies créent des valeurs extrêmes.



Un histogramme peut soutenir nos propos sur le diagramme à moustache, nous remarquons qu'en effet la majorité du temps il ne pleut pas où très peu <2mm (20 000 jours sur environ 27 000 jours).

3. Pression



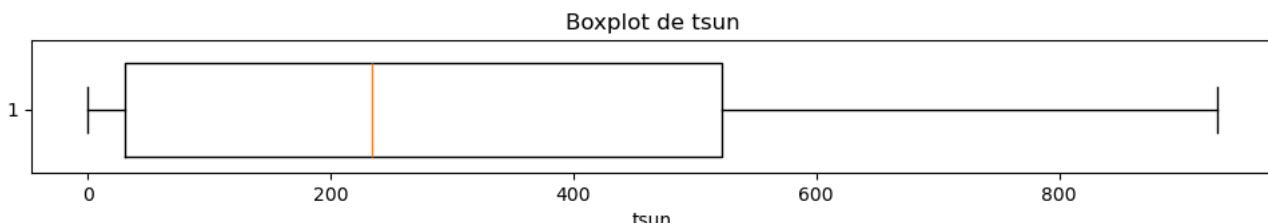
Sur ce diagramme à moustache nous observons que la médiane se situe autour de 1017 hectopascals (hPa), ce qui correspond à la valeur centrale de l'ensemble des relevés.

Le premier quartile (Q1) est approximativement à 1012 hPa, tandis que le troisième quartile (Q3) se situe autour de 1022 hPa, indiquant que 50 % des valeurs journalières sont comprises entre ces deux bornes.

Le minimum et le maximum représentés par les extrémités des moustaches sont respectivement de 980 et 1045 hPa.

On remarque également la présence de valeurs atypiques en dehors des moustaches, ce qui correspond à des chutes ou à des augmentations de pression dues à des événements météorologiques (tempête).

4. Durée d'ensoleillement

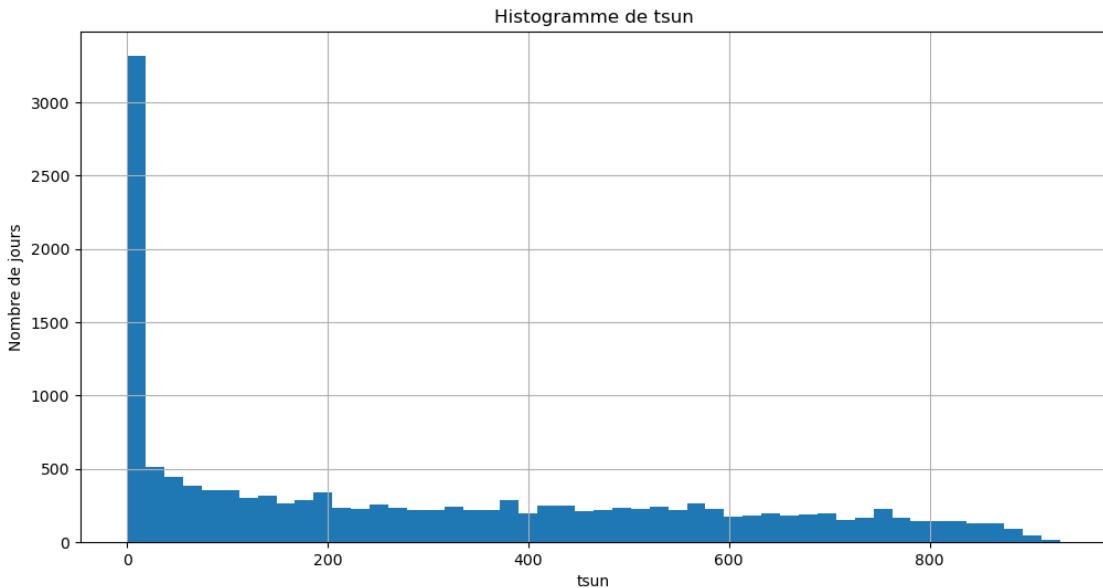


Sur ce diagramme à moustache nous observons que la médiane se situe autour de 300 minutes d'ensoleillement ce qui correspond à la valeur centrale de l'ensemble des relevés.

Le premier quartile (Q1) est approximativement à 30 minutes, tandis que le troisième quartile (Q3) se situe autour de 520 minutes, indiquant que 50 % des valeurs journalières sont comprises entre ces deux bornes.

Le minimum et le maximum représentés par les extrémités des moustaches sont respectivement de 0 et 930 minutes.

Cependant, sur ce diagramme, il est difficile de visualiser clairement la répartition de l'ensoleillement au quotidien. L'histogramme suivant permet de mieux appréhender cette distribution et montre que, la plupart du temps, la durée d'ensoleillement se situe entre 0 et 30 minutes par jour.



Cette analyse statistique permet d'avoir une première vision globale des données, de leur dispersion et des valeurs extrêmes. Elle met aussi en évidence la présence de données manquantes et une forte variabilité pour certaines variables, éléments à prendre en compte pour la suite de l'analyse des cycles.

III. Outils mathématiques utilisés

Le langage Python ainsi que les bibliothèques *Pandas*, *NumPy* et *Matplotlib* ont été utilisés afin de réaliser les graphiques et spectres utilisés pour l'analyse suivante.

Pour l'analyse temporelle, les signaux météorologiques sont étudiés sous forme de séries temporelles journalières. Nous affichons les données sur des périodes spécifiques pour voir l'évolution des grandeurs au fil du temps, et cela permet en l'occurrence de bien visualiser les cycles annuels.

Pour l'analyse fréquentielle, afin d'identifier l'existence de cycles périodiques dans les données, une analyse fréquentielle est réalisée à l'aide de la transformée de Fourier discrète, implémentée par la fonction *fft* de la bibliothèque *NumPy*. Chaque série temporelle est considérée comme un signal échantillonné quotidiennement (fréquence d'échantillonnage égale à 1 jour⁻¹). La transformée de Fourier permet de passer du domaine temporel au domaine fréquentiel et de décomposer le signal en une somme de composantes sinusoïdales de différentes fréquences.

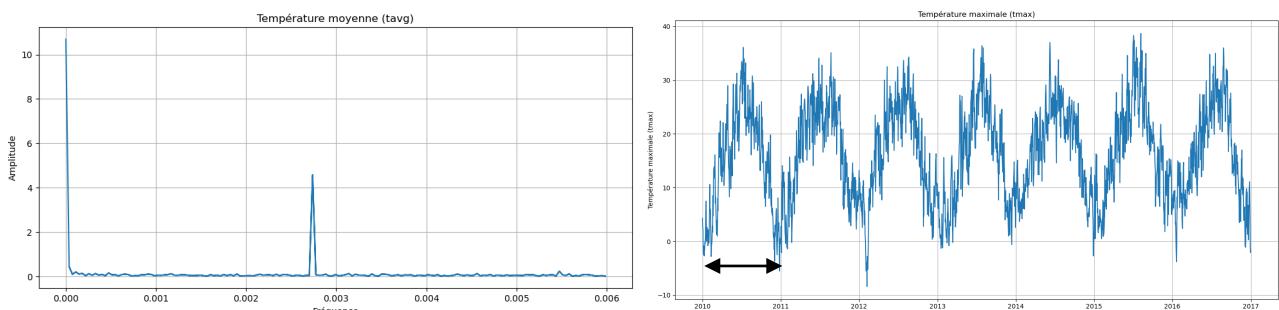
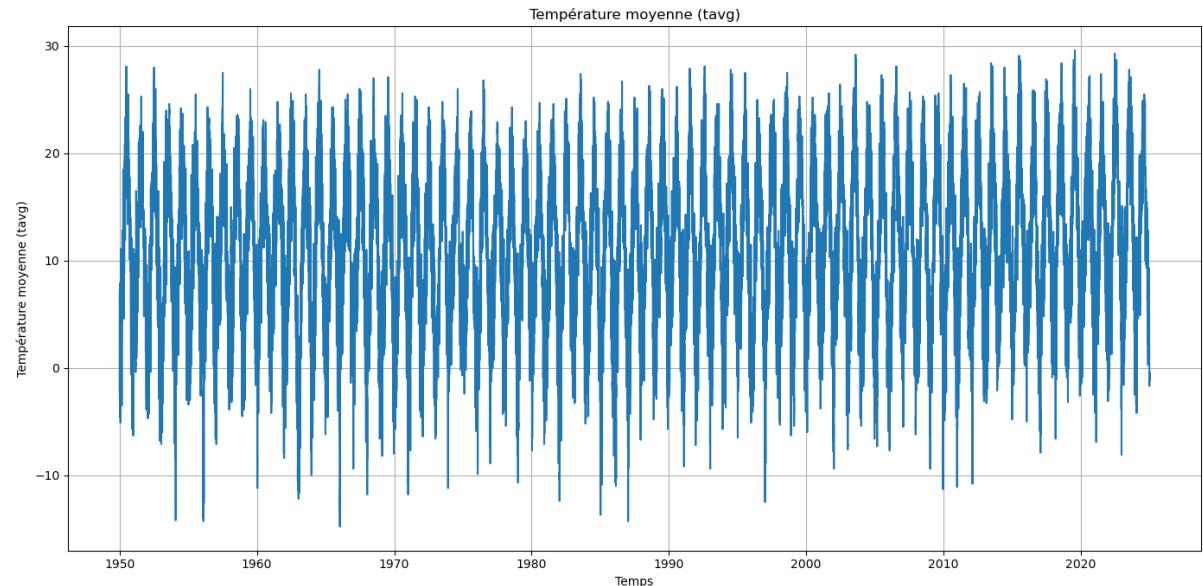
Le spectre d'amplitude est obtenu en prenant le module de la transformée de Fourier et en le normalisant par le nombre de points. Nous utilisons donc ce spectre, en repérant les pics dominants. Chaque pic correspond à une fréquence caractéristique que l'on relève, et qui nous permet de déterminer la période associée grâce à la relation $T = 1/f$.

IV. Cycles annuels, synoptiques et pluri-annuels

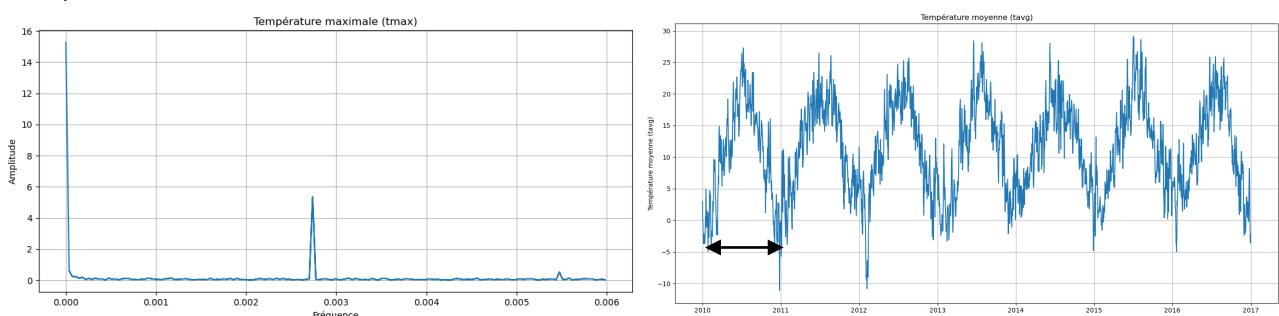
À partir des graphiques temporels et des spectres fréquentiels, nous cherchons donc à identifier les cycles annuels, mais aussi les cycles plus courts (synoptiques) ou plus longs (pluri-annuels). Avec les périodes déterminées nous pouvons déterminer s'il s'agit d'un cycle annuel, synoptique ou pluri-annuel. En particulier, les pics de fréquence environ 0,0027 correspondent aux cycles annuels, ou 365 jours environ.

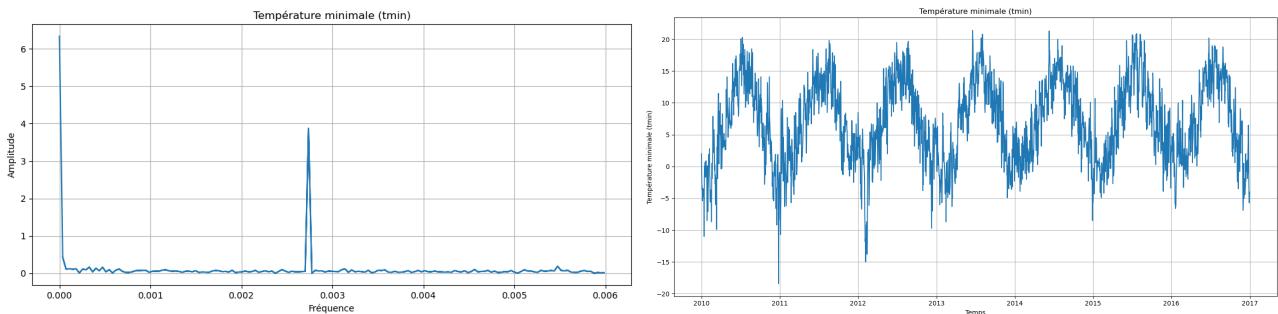
A. Température moyenne, minimale et maximale

Étant donné que ces 3 grandeurs sont liées, nous allons les traiter ensemble. Nous avons des données pour ces 3 grandeurs du 01/01/1950 au 31/12/2024.



La température moyenne est relativement constante. Bien que tous les pics n'aient pas la même amplitude, le bruit généré par ces légères différences ne devrait pas gêner la lecture en fréquentiel (il en va de même pour la température min et max, les 3 graphes se ressemblent fortement, bien qu'on observe une légère baisse des températures au fil du temps pour tmin et tmax).



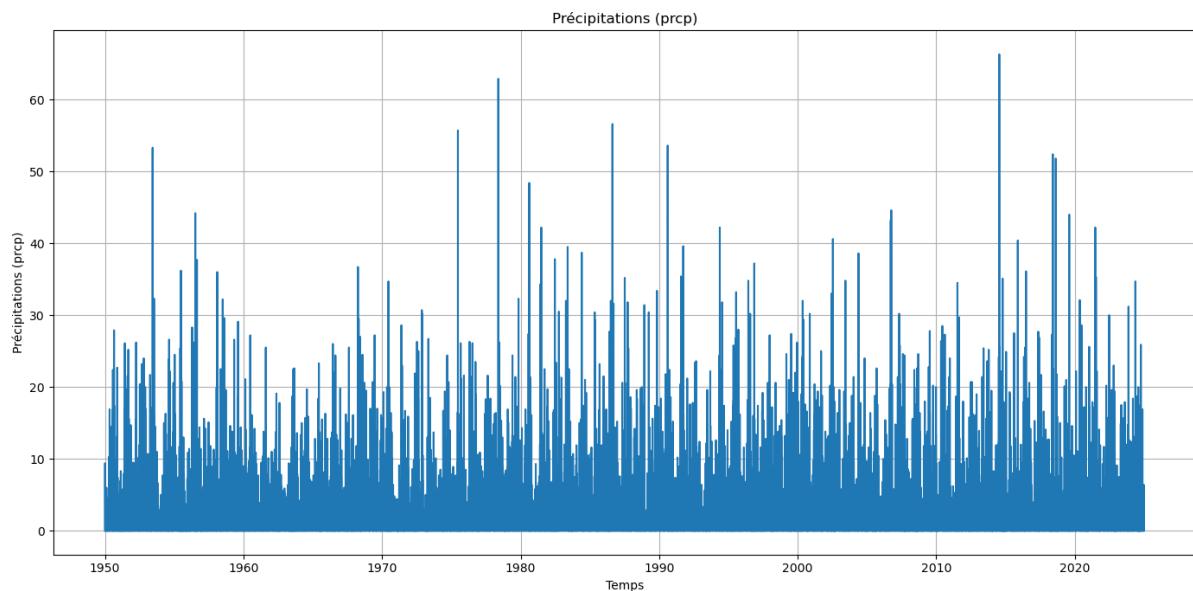


Sur les spectres fréquentiels, nous repérons un léger bruit, mais il n'empêche pas la détermination des différentes périodes. Nous remarquons un pic aux alentours de 0,0027, ce qui correspond à une période annuelle. Cela se confirme avec l'analyse temporelle qui ressemble à un « sinus bruité » avec une valeur maximale aux alentours de juin/juillet/août, qui correspond aux températures les plus hautes de l'année, et une valeur minimale vers décembre/janvier, qui correspond aux températures les plus basses de l'année. Nous pouvons voir que les cycles durent en moyenne 1 an, mais certains cycles peuvent être un peu plus courts, environ 10/11 mois et d'autres un peu plus longs, 13/14, voire 15 mois dans certains cas.

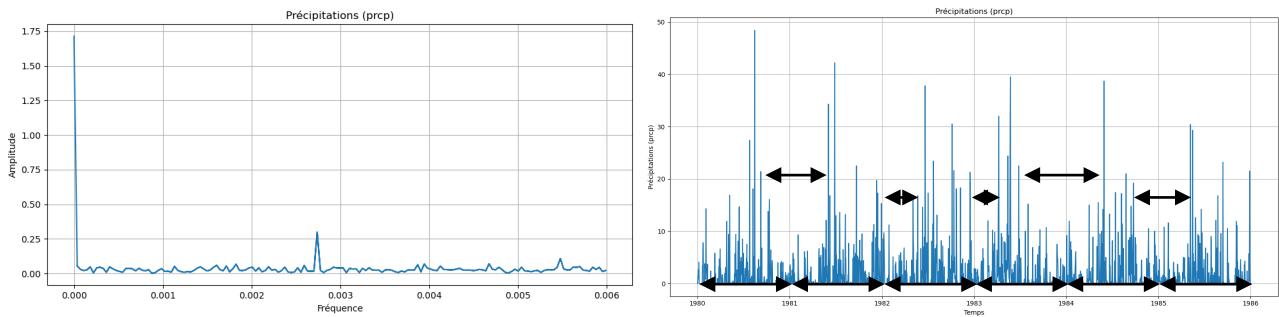
On remarque aussi un petit pic vers 0,0055 qui correspond à une période de 6 mois environ, or ce cycle est sûrement lié au cycle principal annuel montrant la montée (de l'hiver à l'été) et la descente (de l'été à l'hiver) de la température, ce qui génère un signal répétitif tous les 6 mois, donc ce cycle ne peut pas vraiment être considéré comme synoptique.

B. Précipitations

Pour la précipitation, nous avons des données du 01/01/1950 au 31/12/2024.



En ce qui concerne la précipitation, les données sont bien plus hétérogènes que pour les températures. On reconnaît un pattern avec les pics et les creux, mais ceux-ci semblent irréguliers, et cela risque de créer du bruit et de rendre plus difficile la détermination de cycles ayant des pics faibles.

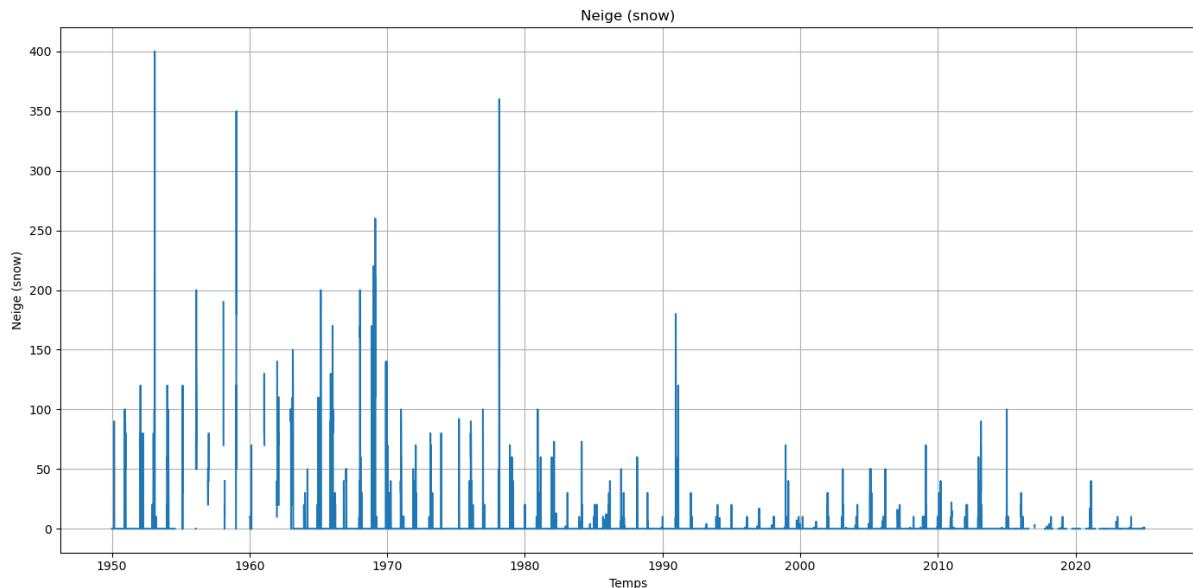


Sur le spectre fréquentiel, nous remarquons toujours le pic à 0,0027 correspondant au cycle annuel. Cela se remarque aussi sur le graphique temporel, où on a un fort pic vers avril/mai et septembre/octobre et un gros creux fin décembre/janvier/février. Cependant, nous remarquons un léger bruit sur le spectre fréquentiel dû à l'irrégularité des précipitations. Le schéma n'est pas exact, le creux se déplace, sa largeur et sa profondeur changent d'une année à l'autre et il en va de même pour les pics. Nous avons ici une période bien plus irrégulière que pour les températures allant de périodes courtes d'environ 8/9 mois à des périodes bien plus longues d'environ 15/16 mois.

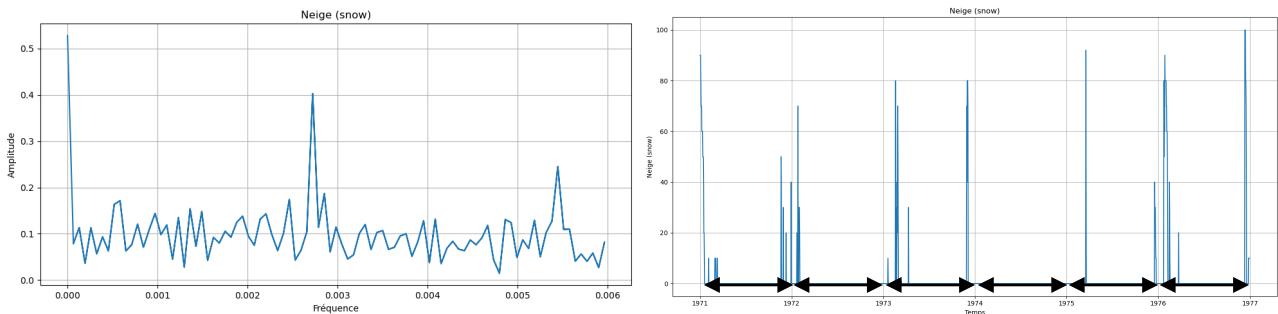
On voit aussi un pic à 0,0055 qui correspond toujours à une période de 6 mois. Cette fois s'il s'agit sûrement d'un cycle synoptique dû au changement de saison humide vers saison sèche (printemps vers été et automne vers hiver). Ce cycle est légèrement visible sur le graphique temporel mais reste bien moins dominant que le cycle annuel.

C. Chutes de neige

Pour les chutes de neige nous avons des données du 01/01/1950 au 31/07/1954 puis du 01/04/1963 au 31/12/2024. Néanmoins, durant ce temps de vide il y avait régulièrement des données non nulles donc nous supposons que les données vides correspondent à des 0.



Les données pour les chutes de neige sont irrégulières, mais pas de la même manière que les précipitations. Les pics sont réguliers en fréquence, mais leur intensité diminue grandement au fil des années, ce qui risque aussi de créer du bruit pour l'analyse fréquentielle.

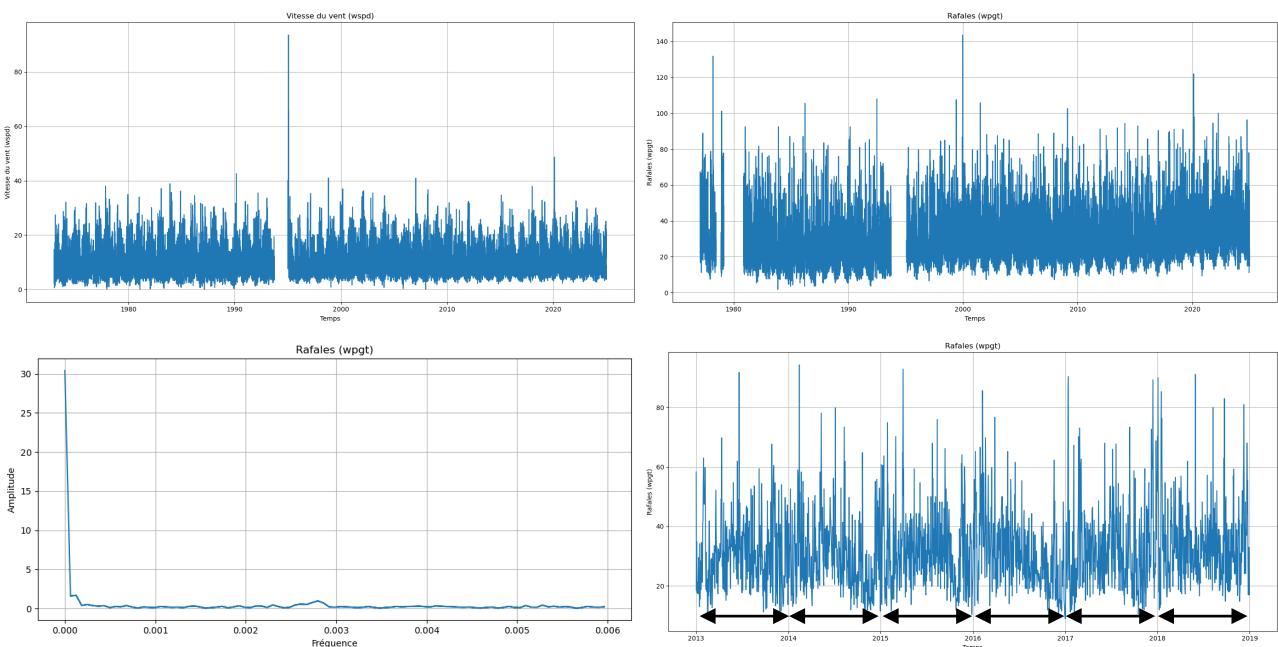


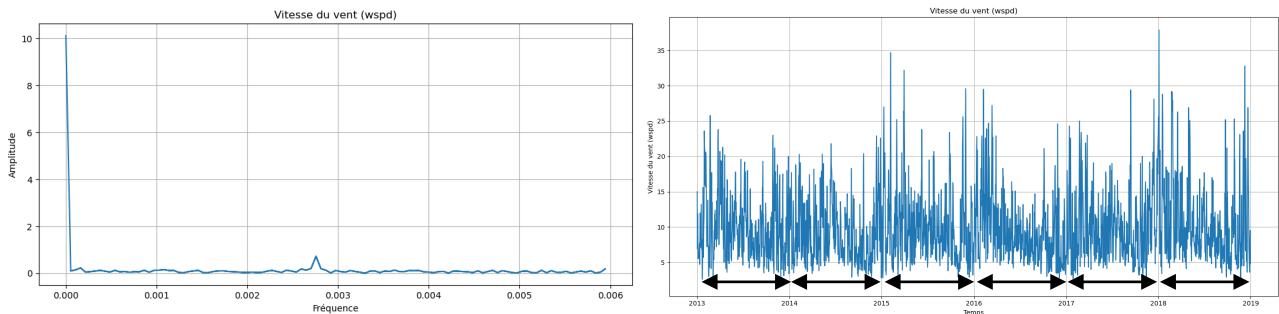
Sur le spectre fréquentiel, nous remarquons toujours le pic à 0,0027 correspondant au cycle annuel. Cela se remarque effectivement sur le graphique temporel où nous voyons des forts pics en hiver. Cependant, les pics ne sont jamais les mêmes mois, allant de janvier à avril. De même, comme nous pouvons le voir, certaines années nous avons une seule chute de neige alors que d'autres nous en avons bien plus. Ce qui fait que la période peut aller de 8/9 mois à 14/15 mois. Cela cause un fort bruit dans le spectre fréquentiel et le rend plus dur à étudier.

Un second pic apparaît autour de la fréquence 0,0055, correspondant à une période d'environ six mois. Toutefois, la neige n'étant pas présente de manière régulière tous les six mois, ce cycle ne peut pas être considéré comme indépendant. Il est vraisemblablement lié au cycle annuel principal et correspond à l'alternance entre les périodes hivernales et printanières, caractérisées par des chutes de neige, et les périodes estivales et automnales, où la neige est absente. Cette alternance induit ainsi une période d'environ six mois.

D. Vitesse du vent et pics de rafales

Pour la vitesse du vent nous avons des données du 01/01/1973 au 30/09/1993 puis du 01/01/1995 au 31/12/2024. Pour les pics de rafale nous avons des données du 27/01/1977 au 31/10/1977 puis du 13/11/1980 au 30/09/1993 puis du 01/02/1995 au 31/12/2024. Comme ces 2 données sont liées (on peut le voir grâce au graphe temporel les pics sont aux mêmes endroits) nous allons les traiter ensemble.

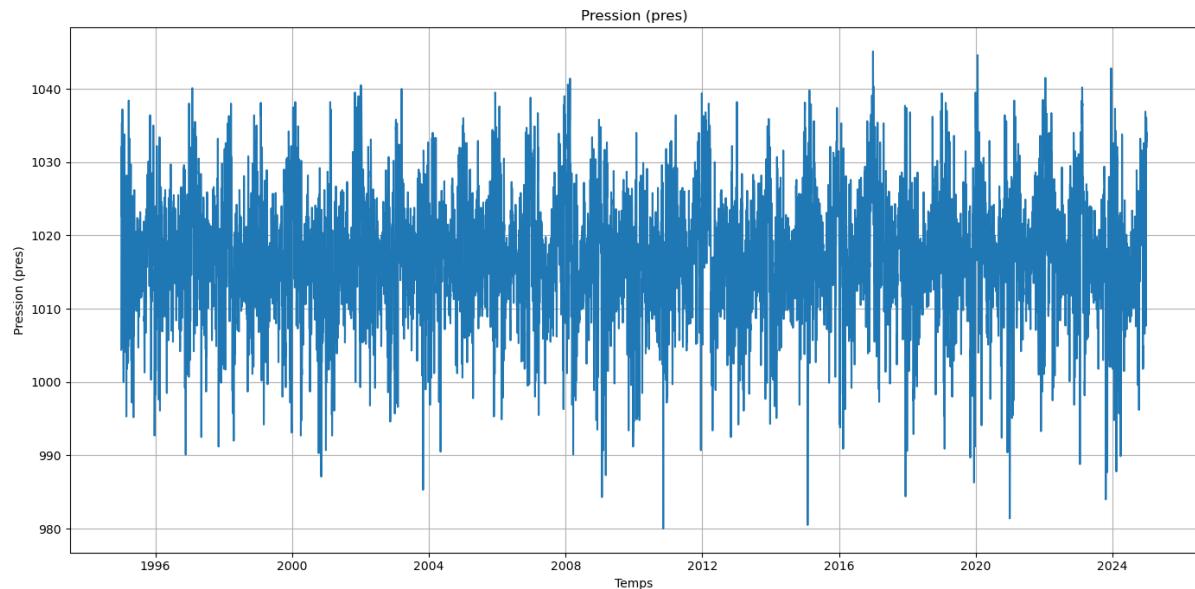




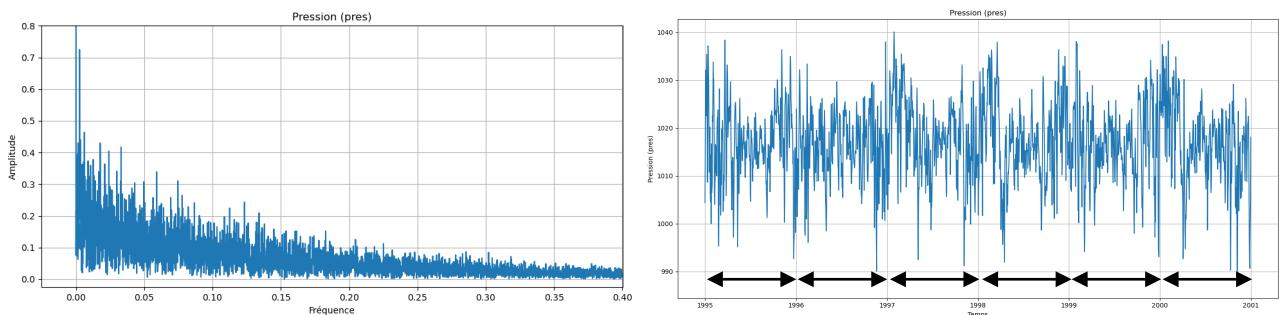
Sur le spectre fréquentiel, nous remarquons un très léger pic à 0,0027 correspondant au cycle annuel. Cela se remarque sur le graphique temporel où nous voyons la vitesse du vent et les rafales plus élevées vers le milieu d'année. Cependant, ces hausses ont parfois lieu en début d'année, vers mars/avril, ce qui fait que certaines périodes durent 9/10 mois et d'autres 13/14 mois, et cela explique le léger bruit que l'on voit sur le spectre fréquentiel.

E. Pression

Nous avons des données sur la pression du 01/01/1995 au 31/12/2024.



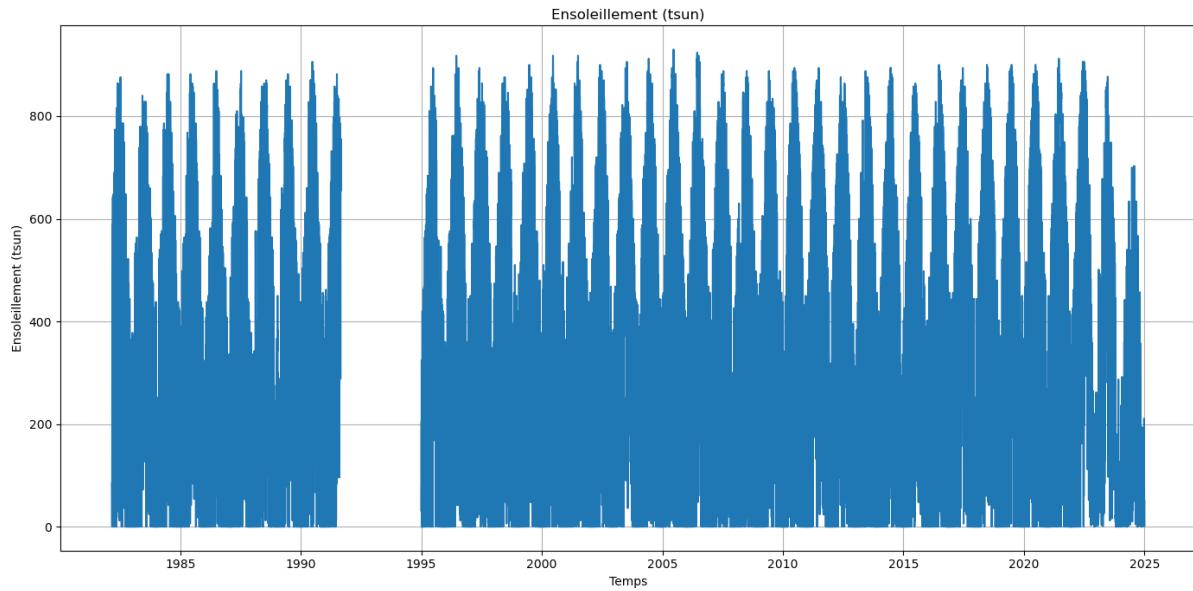
Pour les données sur la pression, nous remarquons que la période et l'amplitude sont assez régulières.



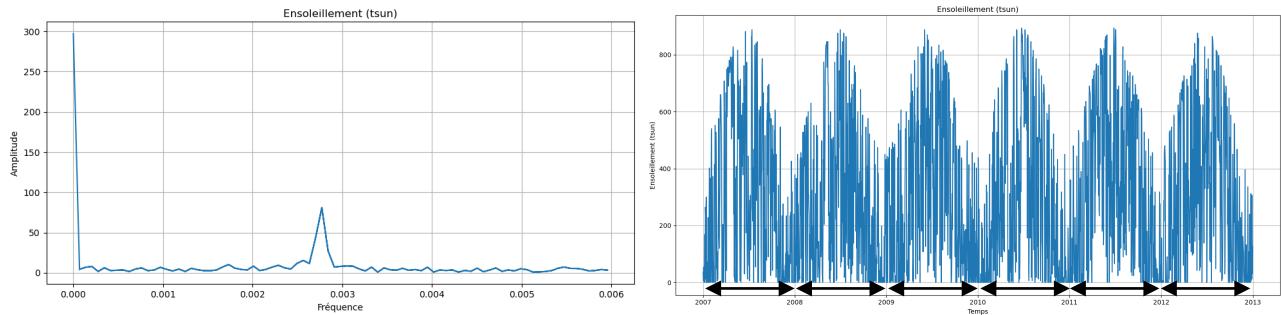
Pour avoir un spectre utilisable, nous avons dû centrer le signal, car sinon le spectre était nul. Cela était dû aux faibles variations de la pression (50 hPa tout au plus) par rapport aux valeurs (en moyenne 1017 hPa). Le spectre ressemble à une composante continue. Cela montre l'absence de cycle périodique strict et met en évidence une variabilité dominée par des phénomènes synoptiques de quelques jours, associés au passage des dépressions et anticyclones. Or, sur le graphique temporel nous voyons bien une périodicité annuelle qui se dessine mais il n'est pas visible sur le spectre fréquentiel car masqué par la forte variabilité synoptique.

F. Ensoleillement

Pour l'ensoleillement nous avons des données du 02/03/1982 au 31/08/1991 puis du 01/01/1995 au 31/12/2024.



Pour les données d'ensoleillement, nous voyons que le signal est constant.



Sur le spectre fréquentiel, nous remarquons toujours le pic à 0,0027, correspondant au cycle annuel. Cela se remarque bien sur le graphique temporel, nous voyons l'ensoleillement augmenter en été puis diminuer en hiver. Les cycles sont globalement de même période, ce qui explique le faible bruit sur le spectre fréquentiel.

V. Conclusion

L'analyse des séries météorologiques de Strasbourg-Entzheim a permis d'identifier des cycles annuels bien marqués pour la plupart des variables : températures, précipitations, neige et ensoleillement. On observe donc que les températures et l'ensoleillement sont maximaux en été et minimaux en hiver, tandis que les chutes de neige apparaissent surtout en hiver.

Des cycles plus courts, d'environ six mois, ont également été repérés pour certaines variables, mais ils ne sont pas indépendants : ils correspondent à l'évolution rapide du signal entre les saisons (passage hiver-été et été-hiver), créant ainsi une répétition semi-annuelle. Par contre, aucun cycle pluri-annuel n'a été détecté, peut-être parce que le climat tempéré continental du lieu n'est pas affecté par les influences océaniques. Il se peut également que des données couvrant une période plus longue auraient été nécessaires pour mettre en évidence des cycles pluriannuels.

Certaines séries présentent un bruit important ou des données manquantes, comme les précipitations ou le vent, ce qui peut masquer des variations faibles. Malgré cela, la périodicité annuelle reste visible dans toutes les séries analysées.