

dd

dd

20 mars 2008

Table des matières

1	Etude descriptive des séries de données	2
1.1	Etude du SO ₂	2
1.2	Etude des données de pluie	3
1.3	Etude du pH	4
1.4	Etude de la conductivité	5
1.5	Etude du Cl ⁻	6
1.6	Etude du NO ₃ ⁻	7
1.7	Etude du SO ₄	8
1.8	Etude du NH ₄ ⁺	9
1.9	Etude du Na ⁺	10
1.10	Etude du K ⁺	11
1.11	Etude du Mg ²⁺	13
1.12	Etude du Ca ₂ ⁺	14
1.13	Etude de la corrélation	16

Chapitre 1

Etude descriptive des séries de données

1.1 Etude du SO2

En ce qui concerne le SO2, il en manque 140 sur 2191, soit un pourcentage de 6,4%.

Par définition, une série chronologique, aussi appelée chronique ou série temporelle est une suite finie de données indexée par le temps. L'indice temps peut-être suivant les cas la minute, l'heure, le jour, le mois, l'année...

Il y a de nombreuses fluctuations de S02, qui peuvent être très importantes d'un jour à l'autre. Parmi elles, on observe 4 pics importants.

Si on divise notre graphique par année, certains semblent logiques car ils ont lieu pendant l'hiver. Toutefois, un d'entre eux se situe en été. D'après les données, il s'agit du 29 juillet 2002 (23,9).

La courbe observée ne présente aucune périodicité.

Min	0,11
1st Qu.	0,36
Median	0,90
Mean	1,55
3rd Qu.	2,02
Max	37,61
NA's	140,00

TAB. 1.1 – informations obtenues grâce à la commande summary

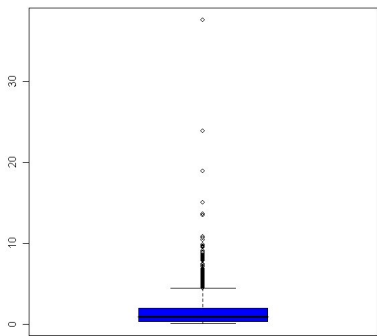


FIG. 1.1 – Boîte à moustache du SO2

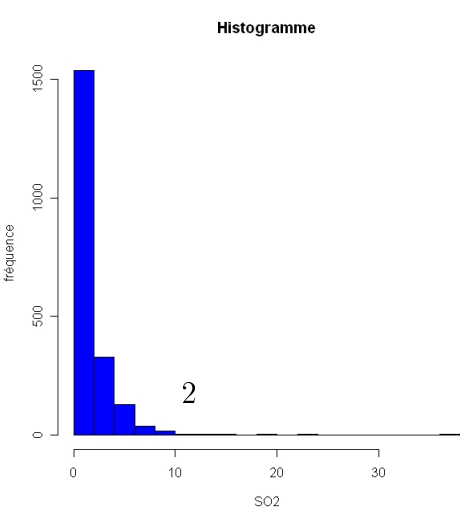


FIG. 1.2 – Histogramme du

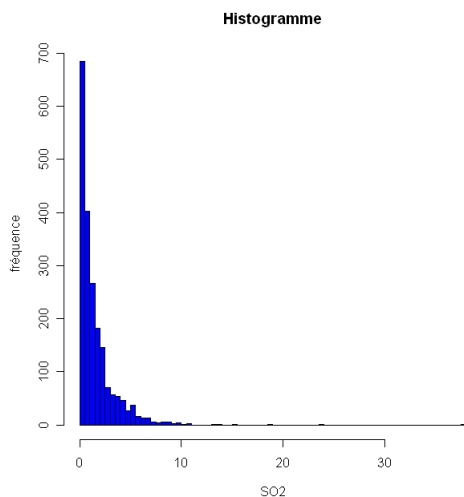


FIG. 1.3 – Histogramme
“plus fin” du SO2

La boîte à moustache montre qu’il y a de nombreuses valeurs aberrantes. On y observe aussi que notre représentation n’est pas symétrique. La moyenne (1,55) est supérieure à la médiane (0,90).

D’après le premier histogramme, la quasi-totalité des données se situe entre 0 et 10 et même que plus de 80 % se situe entre 0 et 5.
Plus du tiers des données se situent entre 0 et 1 d’après le graphique 2.

1.2 Etude des données de pluie

Concernant le relevé des pluies, on constate que sur 2191 données, il y a 1071 jours sans précipitation soit un pourcentage de 48,89%

La courbe ne présente aucune périodicité.

On constate de nombreuses fluctuations.

On observe également 4 pics important dont deux très élevés. Deux des pics ont lieu durant l’hiver ce qui semble logique, cependant les deux autres, ont lieu au printemps et en été. D’après les données, ils se situent le 2 août avec un relevé de 54,5 mm et le 27 avril 2004 avec un relevé de 76 mm.

Min	0,100
1st Qu.	0,500
Median	2,200
Mean	4,489
3rd Qu.	6,000
Max	76,000
NA's	1071,00

TAB. 1.2 – Informations obtenues grâce à la commande summary

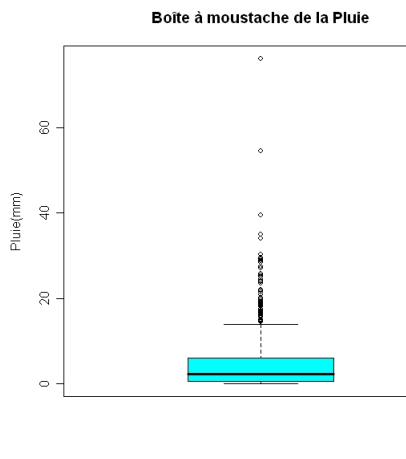


FIG. 1.4 – Boîte à moustache de la quantité de pluie

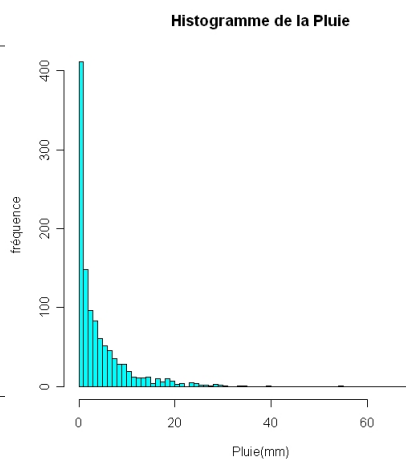


FIG. 1.5 – histogramme du relevé de pluie

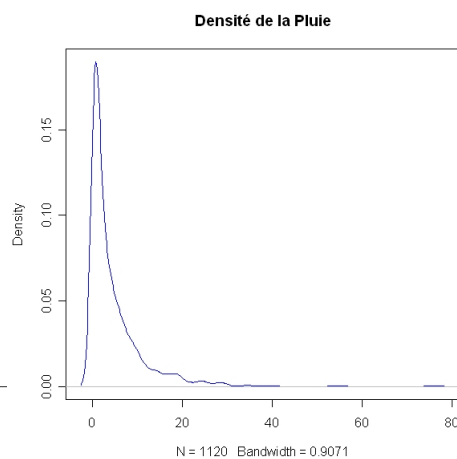


FIG. 1.6 – Densité

La boîte à moustache montre que la médiane (2,200) est inférieure à la moyenne (4,489).

Ceci montre que l'écart entre les valeurs au dessus de la médiane et celle-ci est plus important que celui pour les valeurs en dessous de la médiane. Cette représentation est asymétrique à gauche, et montre qu'il existe de grandes valeurs.

En effet, il y a de nombreuses valeurs atypiques (notamment le 76).

D'après l'histogramme, la quasi-totalité des données se situe entre 0 et 20. Il y a d'ailleurs peu de valeurs supérieures à 20. En regardant de plus près, on constate que plus de 80 % des données se situe entre 0 et 10.

1.3 Etude du pH

Concernant le pH, on constate que sur 2191 données, il en manque 1450 soit un pourcentage de 66,18 %

Par définition :

$$pH = -\log(H^+)$$

De 0 à 7 le pH est acide.

En 7 il est neutre.

De 7 à 14 le pH est basique.

Mais généralement on compare le pH d'une solution avec le pH de l'eau distillée à l'équilibre dans l'air qui est de 5,65. En dessous de 4,5, il y a acidité. Au dessus de 5,65, on parle de solution basique.

La courbe ne présente aucune périodicité.

On constate de nombreuses fluctuations ainsi que de nombreuses discontinuités.

Globalement, on constate que la pluie est acide, le pH se situe presque tout le temps en dessous de 7.

On observe un pic très important dont la valeur est supérieur à 7, ce dernier a lieu en hiver.

Min	3,490
1st Qu.	4,760
Median	5,240
Mean	5,264
3rd Qu.	5,770
Max	7,610
NA's	1450,00

TAB. 1.3 – informations obtenues grâce à la commande summary

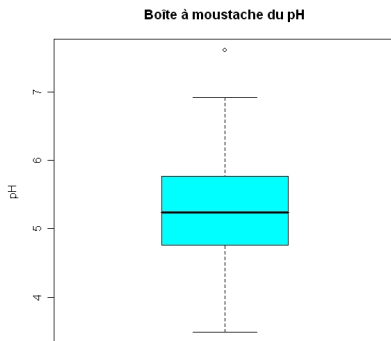


FIG. 1.7 – Boîte à moustache de la valeur du Ph

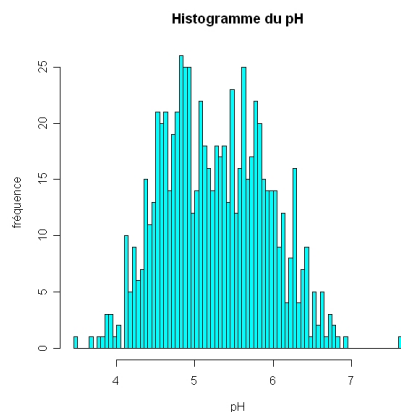


FIG. 1.8 – Histogramme de la valeur du Ph

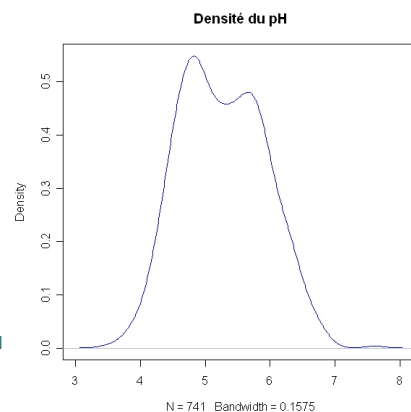


FIG. 1.9 – Densité

D'après la boîte à moustache, la médiane (5,240) est inférieure à la moyenne (5,264). On peut considérer que cette représentation est symétrique. On observe une unique valeur atypique correspondant au cas où le pH est basique (7,61).

L'histogramme montre que les données sont regroupées autour de 5. On constate bien la symétrie vers 5,25. 95 % des données sont inférieures à 7 ce qui montre l'acidité quotidienne de le pluie.

1.4 Etude de la conductivité

Concernant la conductivité, on constate que sur 2191 données, il en manque 1450 soit un pourcentage de 66,18 %. Plus la conductivité est élevée, plus il y a d'ions. L'étude de ce paramètre n'a pas grand intérêt donc nous nous y intéresserons peu par la suite.

La courbe ne présente aucune périodicité. On constate de nombreuses fluctuations ainsi que de nombreuses discontinuités. Il y a 4 pics majeurs situés en début ou fin d'hiver.

Min	1,7
1st Qu.	10,0
Median	17,3
Mean	24,1
3rd Qu.	29,9
Max	257,2
NA's	1450,0

TAB. 1.4 – Informations obtenues grâce à la commande summary

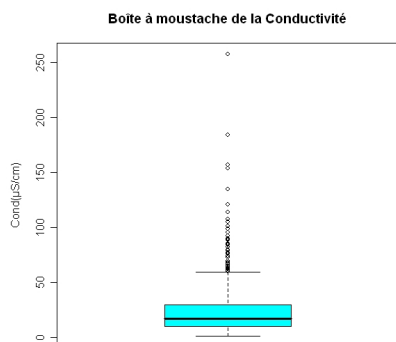


FIG. 1.10 – Boîte à moustache de la conductivité

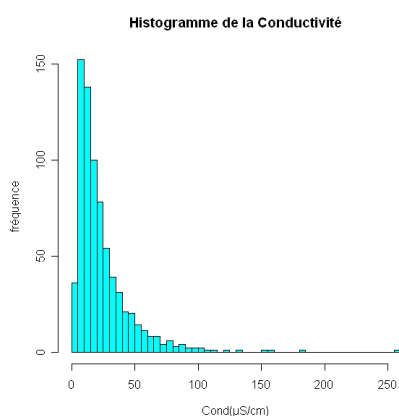


FIG. 1.11 – Histogramme de la conductivité

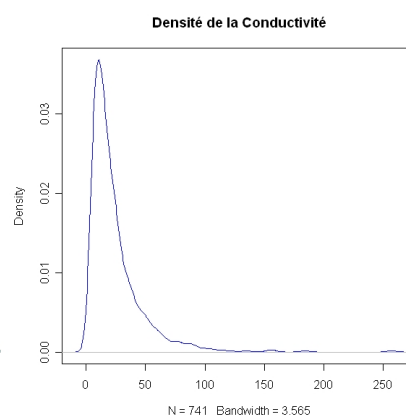


FIG. 1.12 – Densité

La boîte à moustache montre que la médiane (17,3) est inférieure à la moyenne (24,1). Ceci montre que la représentation est asymétrique à gauche.

On observe de nombreuses valeurs atypiques notamment le 257,2 comparé à une moyenne de 24,1.

D'après l'histogramme, la quasi-totalité des données se situe entre 0 et 100. On a plus de 80 % entre 0 et 50.

On observe bien l'asymétrie à gauche ce qui était prévisible d'après la forme de la boîte à moustache. Par ailleurs, les données sont regroupées autour de 5. On constate bien la symétrie vers 5,25.

95 % des données sont inférieures à 7 ce qui montre l'acidité quotidienne de le pluie.

1.5 Etude du Cl^-

Concernant le Cl, on constate que sur 2191 données, il en manque 1479 soit un pourcentage de 67,50 %

On ne remarque pas de périodicité.

Il y a de nombreuses fluctuations et de discontinuités dues aux données manquantes.

On observe 4 pics important se situant en début ou fin d'hiver. Le plus important a lieu en hiver 2005, plus précisément le 18 janvier 2005 avec un relevé de 39,82.

Min	0,030
1st Qu.	0,390
Median	0,980
Mean	2,426
3rd Qu.	2,780
Max	39,820
NA's	1479,0

TAB. 1.5 – informations obtenues grâce à la commande summary

La médiane (0,980) est très inférieure à la moyenne (2,426). Ceci montre qu'il n'y a pas de symétrie dans les données.

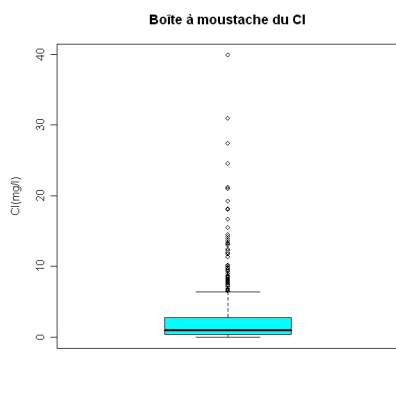


FIG. 1.13 – Boîte à moustache de la concentration en chlore

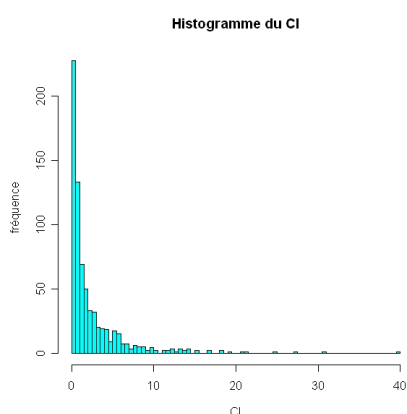


FIG. 1.14 – Histogramme de la concentration en chlore

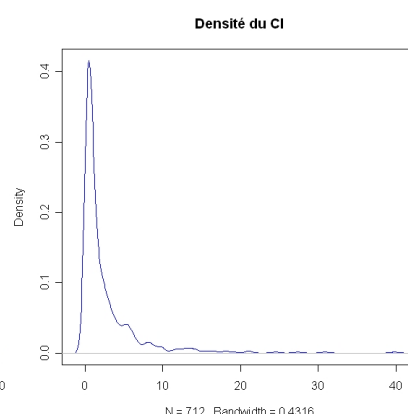


FIG. 1.15 – Densité

La boîte à moustache est très "condensée", on constate de nombreuses valeurs atypiques.

La représentation est asymétrique à gauche et on peut observer de grandes valeurs.

On observe bien l'asymétrie à gauche sur l'histogramme. De plus, la globalité des données se situent entre 0 et 10. On observe très peu de valeurs supérieures à 20. Plus de 80 % des données se trouvent entre 0 et 5.

1.6 Etude du NO₃-

Concernant le NO₃⁻, on constate que sur 2191 données, il en manque 1479 soit un pourcentage de 67,50 %

La courbe ne présente aucune périodicité.

Il y a de nombreuses fluctuations, celles-ci ne dépassent pas 4 sauf pour une valeur extrême de 9,5800.

On observe 3 pics principales ainsi que d'autres pics plus faible se situant de façon inexplicable : parfois en hiver, été, ou printemps.

Min	0,0100
1st Qu.	0,1300
Median	0,2400
Mean	0,3949
3rd Qu.	0,4625
Max	9,5800
NA's	1479,0

TAB. 1.6 – informations obtenues grâce à la commande summary

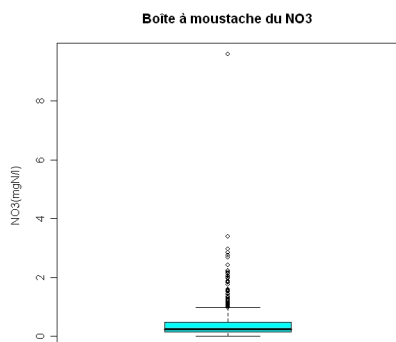


FIG. 1.16 – Boîte à moustache de la concentration en NO_3^-

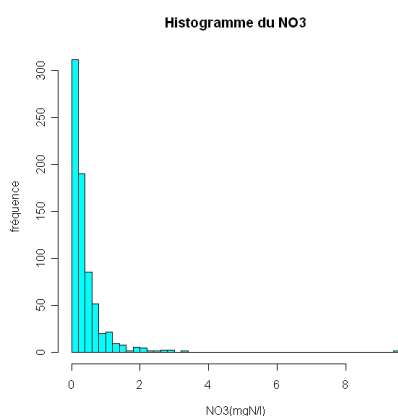


FIG. 1.17 – histogramme de la concentration en NO_3^-

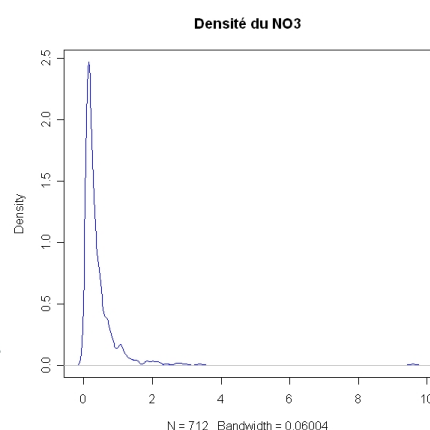


FIG. 1.18 – densité

Dans la boîte à moustache, on observe que la médiane (0,2400) est inférieure à la moyenne (0,3949).

Ceci montre que cette représentation est asymétrique à gauche. On observe également de nombreuses valeurs atypiques.

L'asymétrie est bien visible sur l'histogramme. La globalité des données se situent entre 0 et 2.

Plus de 80 % des données se trouve entre 0 et 1 ce qui explique que la moyenne soit inférieure à 1.

1.7 Etude du SO4

Concernant le SO4, on constate que sur 2191 données, il en manque 1479 soit un pourcentage de 67,50 %

La courbe ne présente aucune périodicité.

On observe de nombreuses fluctuations et discontinuités.

On peut considérer qu'on a 4 pics importants se déroulant en hiver ou au printemps (deux sont très proches de 3).

Les données relevées ne dépassent pas 4, sauf le 26 juin 2003 où le relevé est de 8,070, donnée qui semble totalement aberrante.

Min	0,040
1st Qu.	0,250
Median	0,460
Mean	0,629
3rd Qu.	0,820
Max	8,070
NA's	1479,0

TAB. 1.7 – informations obtenues grâce à la commande summary

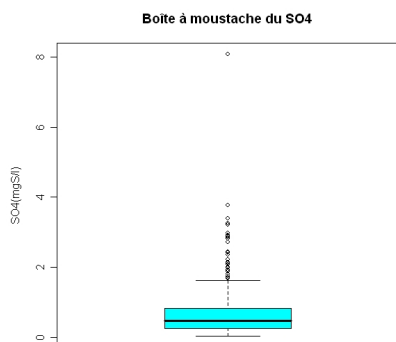


FIG. 1.19 – représentation

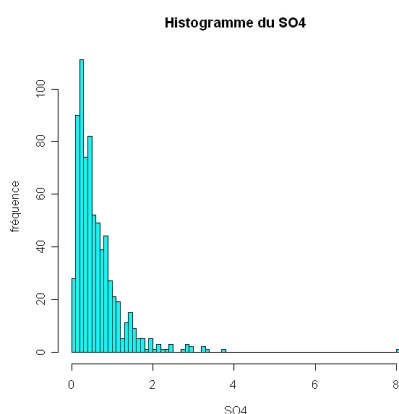


FIG. 1.20 – représentation
du taux de SO4 en fonction
du temps

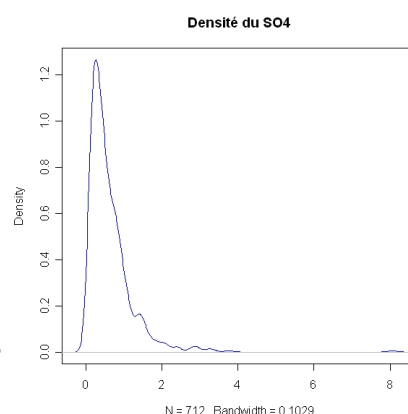


FIG. 1.21 – densité

Dans la boîte à moustache, la médiane (0,460) est inférieure à la moyenne (0,629). Ceci montre que la représentation est asymétrique à gauche. On constate de nombreuses valeurs atypiques ne dépassant pas 4 et une unique valeur supérieure à 8.

Sur l'histogramme, on constate bien la non symétrie des données. La majorité des données se situent entre 0 et 2. On constate même que plus de 80 % des données se situent entre 0 et 1. Une seule valeur dépasse 4. Et il y a très peu de valeurs entre 2 et 4.

1.8 Etude du NH_4^+

En ce qui concerne le NH_4^+ , il manque 1476 données sur 2191.

Il y a des fluctuations de NH_4^+ . Celles-ci peuvent être très importantes d'un jour à l'autre.

On observe des pics importants (environ 3) répartis en hiver le plus souvent.

On n'observe pas vraiment de périodicité sur les 6 années.

Min	0,0100
1st Qu.	0,1900
Median	0,3500
Mean	0,5212
3rd Qu.	0,6400
Max	7,5800
NA's	1476,000

TAB. 1.8 – informations obtenues grâce à la commande summary

On remarque que la moyenne (0,5212) et la médiane (0,3500) ont un écart important, ce qui signifie que notre histogramme et notre boîte à moustache ne seront pas symétriques.

Comme la moyenne est supérieure à la médiane, la boîte à moustache et l'histogramme seront asymétriques à gauche à cause des valeurs aberrantes.

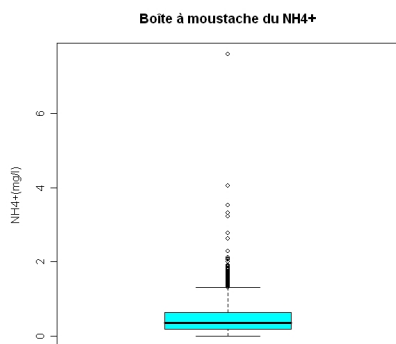


FIG. 1.22 – Boîte à moustache

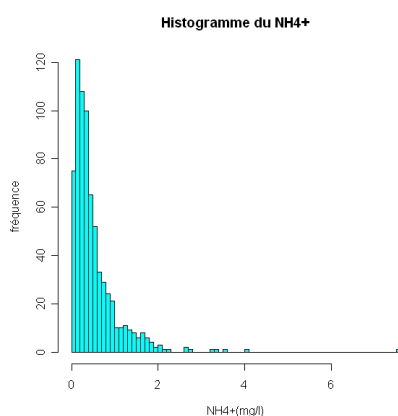


FIG. 1.23 – Histogramme de NH_4^+

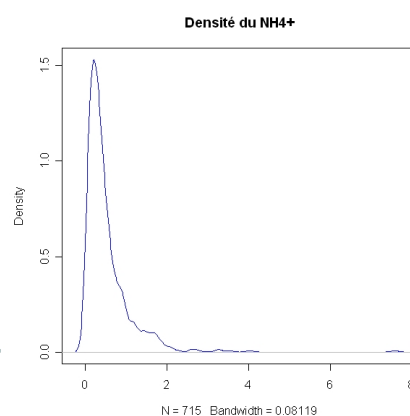


FIG. 1.24 – Densité

Il y a un nombre conséquent de valeurs aberrantes. On observe aussi que notre représentation n'est pas symétrique. La moyenne étant supérieure à la médiane, la boîte à moustache est asymétrique à gauche.

Sur l'histogramme, on voit que sur les 712 données que nous avons en notre possession, environ 400 se situent entre 0 et 0,4 ce qui explique que la médiane soit inférieure à 0,4.

La moyenne est plus élevée (0,5212) à cause des valeurs aberrantes.

1.9 Etude du Na_+

En ce qui concerne le Na_+ , il manque 1479 données sur 2191.

Il y a des fluctuations de Na_+ . Celles-ci peuvent être très importantes d'un jour à l'autre.

On observe des pics importants (environ 3) répartis en hiver le plus souvent.

On n'observe pas vraiment de périodicité sur les 6 années.

Min	0,010
1st Qu.	0,240
Median	0,610
Mean	1,408
3rd Qu.	1,632
Max	21,850
NA's	1479,000

TAB. 1.9 – Informations obtenues grâce à la commande summary

On remarque que la moyenne (0,610) et la médiane (1,408) ont un écart important, ce qui signifie que notre histogramme et notre boîte à moustache ne seront pas symétriques. Comme la moyenne est supérieure à la médiane, la boîte à moustache et l'histogramme seront asymétriques à gauche à cause des valeurs aberrantes.

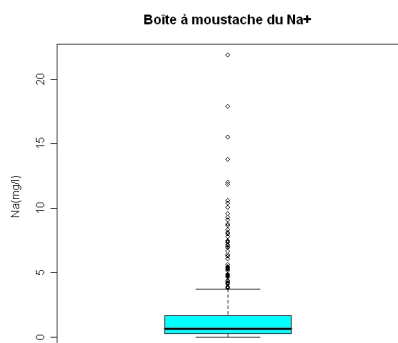


FIG. 1.25 – Boîte à moustache

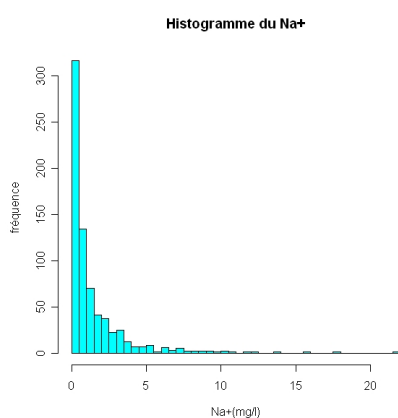


FIG. 1.26 – Histogramme de Na_+

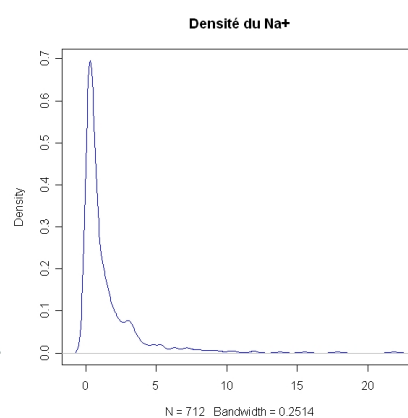


FIG. 1.27 – Densité

Il y a un nombre conséquent de valeurs aberrantes. On observe aussi que notre représentation n'est pas symétrique. La moyenne étant supérieure à la médiane, la boîte à moustache est asymétrique à gauche.

Sur les 712 données que nous avons en notre possession, 450 données se situent entre 0 et 1 (et 550 entre 0 et 2), ce qui explique que notre médiane soit bien inférieure à 1.

1.10 Etude du K_+

En ce qui concerne le K_+ , il manque 1479 données sur 2191.

Il y a des fluctuations de K_+ . Celles-ci peuvent être très importantes d'un jour à l'autre.

On observe des pics importants (environ 6) répartis en début d'hiver le plus souvent, mais pas uniquement.

On n'observe pas vraiment de périodicité sur les 6 années.

Min	0,0100
1st Qu.	0,0300
Median	0,0700
Mean	0,1253
3rd Qu.	0,1400
Max	1,5900
NA's	1479,0000

TAB. 1.10 – Informations obtenues grâce à la commande summary

On remarque que la moyenne (0,1253) et la médiane (0,0700) ne sont pas proches, ce qui signifie que notre histogramme et notre boîte à moustache ne seront pas symétriques. Comme la moyenne est supérieure à la médiane, la boîte à moustache et l'histogramme seront asymétriques à gauche à cause des valeurs aberrantes.

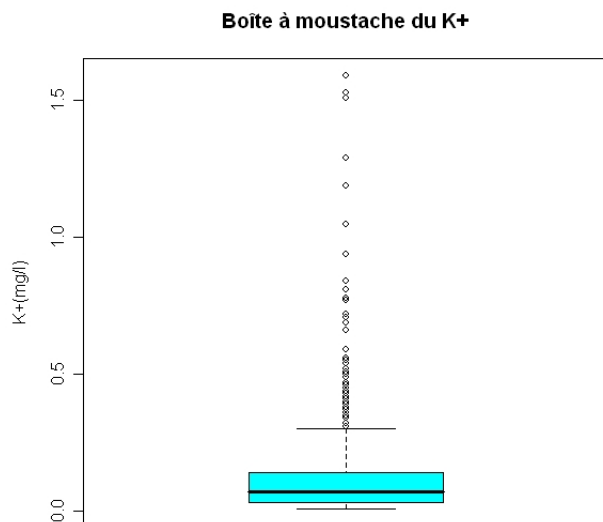


FIG. 1.28 – Boîte à moustache

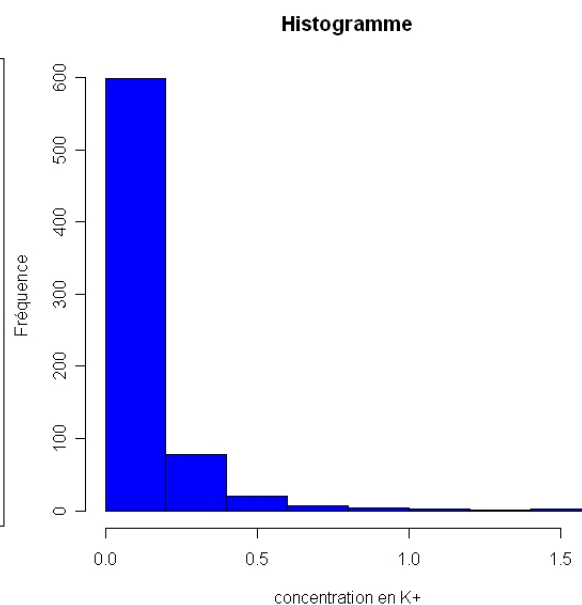


FIG. 1.29 – Histogramme de K_+

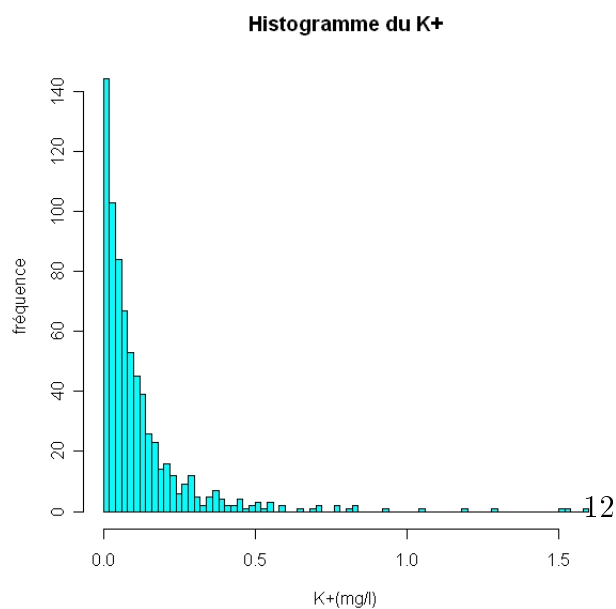


FIG. 1.30 – Histogramme plus précis

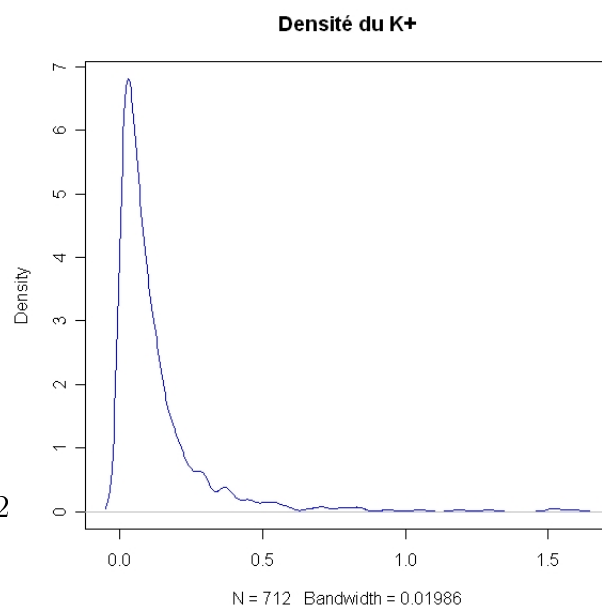


FIG. 1.31 – Densité

Il y a un certain nombre de valeurs aberrantes. On observe aussi que notre représentation n'est pas symétrique. La moyenne étant supérieure à la médiane, la boîte à moustache est asymétrique à gauche.

Sur les 712 données que nous avons en notre possession, 600 se situent entre 0 et 0,2 ce qui explique que la médiane soit inférieure à 0,2. Grâce au deuxième histogramme, nous pouvons être plus précises. 450 données se situent entre 0 et 0,1, ce qui explique que notre médiane soit bien inférieure à 0,4 et notre moyenne également. D'après cet histogramme, il est logique que notre médiane soit de inférieure à 0,1.

1.11 Etude du Mg^{2+}

En ce qui concerne le Mg^{2+} , il manque également 1479 données sur 2191.

Il y a des fluctuations de Mg^{2+} , qui peuvent être très importantes d'un jour à l'autre. Parmi elles, on observe 3 pics très importants et plusieurs autres importants. La majorité des pics se produisent pendant la période hivernale.

Il n'y a pas de périodicité sur les 6 années.

Min	0,0100
1st Qu.	0,0500
Median	0,1100
Mean	0,2013
3rd Qu.	0,2500
Max	2,6700
NA's	1479,0000

TAB. 1.11 – informations obtenues grâce à la commande summary

On remarque que la moyenne et la médiane ne sont pas proche, ce qui signifie que notre histogramme et notre boîte à moustache ne seront pas symétriques.

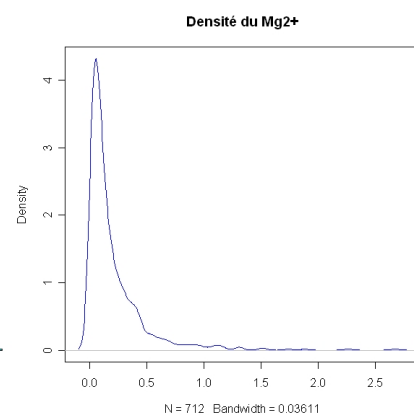
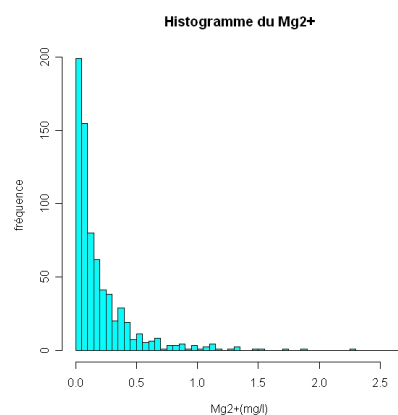
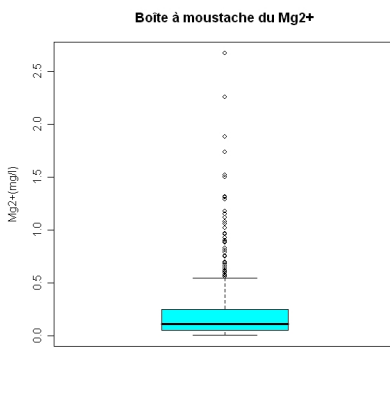


FIG. 1.32 – Boîte à moustache

FIG. 1.33 – Histogramme

FIG. 1.34 – Densité

Il y a un certain nombre de valeurs aberrantes. On observe aussi que notre représentation n'est pas symétrique. La moyenne étant supérieure à la médiane, la boîte à moustache est asymétrique à gauche.

Sur les 712 données que nous avons en notre possession, 500 se situent entre 0 et 0,2 ce qui expliquent que la moyenne soit aux alentours de 0,2 et que la médiane soit inférieure à 0,2.

Comme prévu, on voit bien que l'histogramme est asymétrique à gauche.

1.12 Etude du Ca_2^+

On observe aucune périodicité sur les 6 années.

Min	0,0100
1st Qu.	0,1100
Median	0,2000
Mean	0,3431
3rd Qu.	0,3600
Max	5,3100
NA's	1479,0000

TAB. 1.12 – Informations obtenues grâce à la commande summary

On remarque que la moyenne et la médiane ne sont pas proches, ce qui signifie que notre histogramme et notre boîte à moustache ne seront pas symétriques.

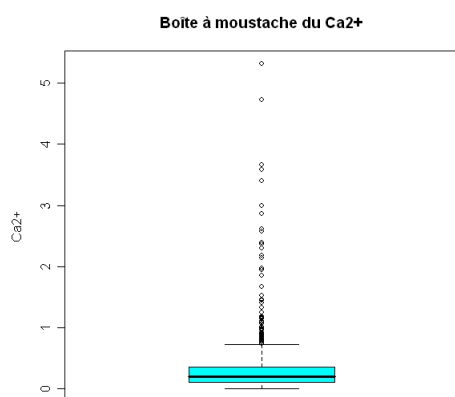


FIG. 1.35 – Boîte à moustache

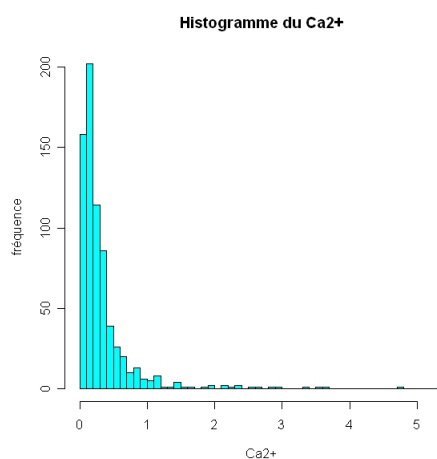


FIG. 1.36 – Histogramme plus précis

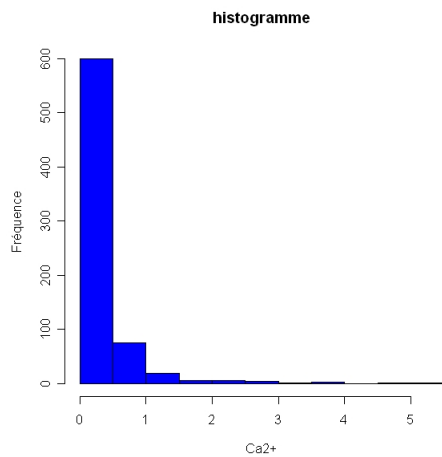


FIG. 1.37 – Histogramme

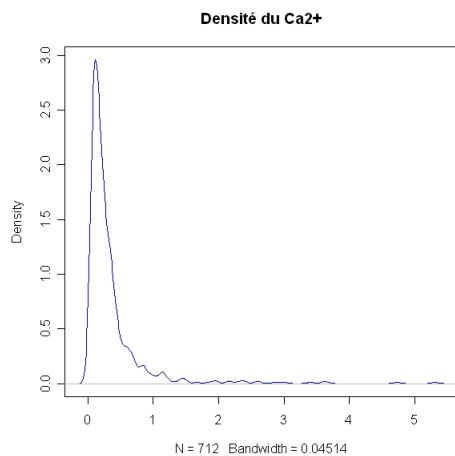


FIG. 1.38 – Densité

Il y a un certain nombre de valeurs aberrantes. On observe aussi que notre représentation n'est pas symétrique.

La moyenne étant supérieure à la médiane, la boîte à moustache est asymétrique à gauche.

D'après le premier histogramme, sur les 712 données que nous avons en notre possession, 600 se situent entre 0 et 0,5 ce qui explique que la médiane soit inférieure à 0,5. Grâce au deuxième histogramme, nous pouvons être plus précises. Un peu moins de 600 données se situent entre 0 et 0,4, ce qui explique que notre médiane soit bien inférieure à 0,4 et notre moyenne également.

Comme environ 375 données se situent entre 0 et 0,2, il est logique que notre médiane soit de 0,2. Comme il y a quelques valeurs atypiques, notre moyenne est donc supérieure à la médiane.