

Validation et correction des séries temporelles d'états de mer et de niveaux d'eau

Sommaire

| | |
|--|-----------|
| 4.1 Outils statistiques pour la comparaison de séries temporelles . . . | 36 |
| 4.1.1 Comparaison de deux séries temporelles à dates communes | 36 |
| 4.1.2 Comparaison de deux séries temporelles à dates non nécessairement communes | 37 |
| 4.1.3 Correction d'une série temporelle à partir d'une autre | 38 |
| 4.2 Comparaisons des simulations avec des observations <i>in situ</i> . . . | 39 |
| 4.2.1 Au point de la bouée d'Anglet | 39 |
| 4.2.2 Au point de la bouée de Saint-Jean-de-Luz | 44 |
| 4.2.3 Au point du marégraphe de Socoa | 45 |
| 4.3 Séries temporelles corrigées en distribution | 47 |

Dans le chapitre 2, nous avons introduits les variables, appelées *variables d'aléa*, pertinentes pour étudier les évènements de submersion. Dans le chapitre 3, les différents jeux de données océaniques disponibles pour l'étude considérée ont été présentés. Avant de discuter, dans le chapitre suivant, de dégâts à la côte que les évènements de submersion peuvent occasionner, des outils de validation et de correction des données seront décrits dans ce chapitre. Ces outils seront ensuite utilisés pour valider ou corriger les séries temporelles d'états de mer et de niveaux d'eau décrites dans le chapitre précédent. L'étape de validation consiste à vérifier que les séries temporelles disponibles (essentiellement issues de simulations numériques) de variables comme H_s , T_p , η , . . . , ont des caractéristiques statistiques proches de séries d'observations *in situ*. En cas de biais important, nous expliquerons comment procéder à une correction fondée sur une version empirique de la *transformation en probabilité* et de la *transformation quantile*.

La première section de ce chapitre est consacrée à la présentation des outils statistiques nécessaires à la validation et l'éventuelle correction des séries tempo-

relles. La section suivante décrit l'utilisation de ces outils pour comparer les séries issues de simulations numériques avec des séries d'observations *in situ*. Enfin, la dernière section consiste en une présentation des séries temporelles de variables d'aléa éventuellement corrigées retenues pour l'étude.

Notons que le contenu de ce chapitre a donné lieu à une publication dans le *Journal of Coastal Research* lors de la conférence "International Coastal Symposium" (ICS 2018) à Busan, Corée du Sud, en mai 2018. L'article publié est disponible en annexe pages 112 à 118.

4.1 Outils statistiques pour la comparaison de séries temporelles

Nous considérons dans cette partie deux séries temporelles univariées $(x_t)_{t \in \mathcal{T}_1}$ et $(y_t)_{t \in \mathcal{T}_2}$ qui peuvent, par exemple, respectivement correspondre à des observations *in situ* et à des simulations numériques au même lieu. Dans la première section, nous expliquons comment ces deux séries peuvent être comparées lorsque les ensembles de dates \mathcal{T}_1 et \mathcal{T}_2 sont identiques. Dans la deuxième section, nous nous intéressons au cas plus général où \mathcal{T}_1 et \mathcal{T}_2 ne sont pas nécessairement identiques. Enfin, dans la dernière section, nous présentons une méthode de correction d'une série temporelle à partir d'une autre afin que la première puisse récupérer certaines caractéristiques statistiques de la seconde. Le lecteur pourra se rapporter au livre de Saporta [32] pour plus de détails.

4.1.1 Comparaison de deux séries temporelles à dates communes

L'outil le plus simple pour comparer $(x_t)_{t \in \mathcal{T}_1}$ et $(y_t)_{t \in \mathcal{T}_2}$ lorsque $\mathcal{T}_1 = \mathcal{T}_2$ est graphique. Il s'agit du nuage de points $\{(x_t, y_t) : t \in \mathcal{T}_1\}$. Plus les points (x_t, y_t) , $t \in \mathcal{T}_1$, sont concentrés autour de la droite d'équation $y = x$, plus les deux séries temporelles sont similaires.

Cette similarité peut également être évaluée à l'aide de plusieurs indicateurs numériques. Notons $\mathcal{T}_1 = \{t_1, \dots, t_n\}$. La différence, en moyenne, entre $(x_t)_{t \in \mathcal{T}_1}$ et $(y_t)_{t \in \mathcal{T}_1}$, peut être évaluée par la formule

$$\text{Biais}[(x_t)_{t \in \mathcal{T}_1}, (y_t)_{t \in \mathcal{T}_1}] = \frac{1}{n} \sum_{t=t_1}^{t_n} (x_t - y_t),$$

laquelle, lorsque $(x_t)_{t \in \mathcal{T}_1}$ représente des observations *in situ* et $(y_t)_{t \in \mathcal{T}_1}$ des simulations correspondantes, peut s'interpréter comme le *biais* de simulation, c'est-à-dire, l'erreur moyenne faite par la simulation considérée. Cette quantité mesure la tendance globale des valeurs simulées à sur- ou sous-estimer la grandeur réelle. Une valeur de zéro (resp. positive, négative) s'interprète comme l'absence d'erreur en moyenne (resp. une sous-estimation, sur-estimation) de la réalité, en moyenne.

Il est également possible de normaliser le biais en calculant le biais en pourcentage comme suit :

$$\text{Biais}_{\%}[(x_t)_{t \in \mathcal{T}_1}, (y_t)_{t \in \mathcal{T}_1}] = \frac{\text{Biais}[(x_t)_{t \in \mathcal{T}_1}, (y_t)_{t \in \mathcal{T}_1}]}{\frac{1}{n} \sum_{t=t_1}^{t_n} x_t} \times 100 = \frac{\bar{x} - \bar{y}}{\bar{x}} \times 100.$$

L'erreur de simulation peut être évaluée par la racine de l'erreur quadratique moyenne :

$$\text{RMSE}[(x_t)_{t \in \mathcal{T}_1}, (y_t)_{t \in \mathcal{T}_1}] = \sqrt{\frac{1}{n} \sum_{t=t_1}^{t_n} (x_t - y_t)^2}.$$

C'est une mesure de précision mélangeant l'erreur moyenne et la variation au sein des erreurs individuelles. Notons que cette mesure est plutôt sensible aux valeurs aberrantes.

En complément des indicateurs précédents, il est d'usage de calculer également le coefficient de détermination :

$$R^2[(x_t)_{t \in \mathcal{T}_1}, (y_t)_{t \in \mathcal{T}_1}] = \text{Corr}[(x_t)_{t \in \mathcal{T}_1}, (y_t)_{t \in \mathcal{T}_1}]^2$$

où $\text{Corr}[(x_t)_{t \in \mathcal{T}_1}, (y_t)_{t \in \mathcal{T}_1}]$ est le coefficient de corrélation linéaire entre les deux séries. Cette quantité mesure la force de l'association linéaire entre $(x_t)_{t \in \mathcal{T}_1}$ et $(y_t)_{t \in \mathcal{T}_1}$. Plus les points (x_t, y_t) , $t \in \mathcal{T}_1$, se concentrent sur une droite (pas nécessairement, la droite d'équation $y = x$), plus le R^2 sera proche de 1. En cas de biais nul, le R^2 pourra être interprété comme le degré de concentration autour de la droite d'équation $y = x$.

4.1.2 Comparaison de deux séries temporelles à dates non nécessairement communes

Dans le cas où les ensembles de dates \mathcal{T}_1 et \mathcal{T}_2 ne sont pas nécessairement identiques, les deux séries $(x_t)_{t \in \mathcal{T}_1}$ et $(y_t)_{t \in \mathcal{T}_2}$ peuvent être comparées en terme de distribution. En supposant que $(x_t)_{t \in \mathcal{T}_1}$ et $(y_t)_{t \in \mathcal{T}_2}$ sont des réalisations de séries temporelles *stationnaires*, une première approche consiste à comparer les fonctions de répartitions empiriques correspondantes :

$$\hat{F}_x(z) = \frac{1}{|\mathcal{T}_1|} \sum_{t \in \mathcal{T}_1} \mathbf{1}(x_t \leq z) \quad \text{et} \quad \hat{F}_y(z) = \frac{1}{|\mathcal{T}_2|} \sum_{t \in \mathcal{T}_2} \mathbf{1}(y_t \leq z), \quad z \in \mathbb{R}, \quad (4.1)$$

où $|\mathcal{T}_1|$ et $|\mathcal{T}_2|$ désignent les cardinaux des ensembles \mathcal{T}_1 et \mathcal{T}_2 , respectivement. Pour tout $z \in \mathbb{R}$, $\hat{F}_x(z)$ (resp. $\hat{F}_y(z)$) correspond simplement à la proportion de x_t , $t \in \mathcal{T}_1$ (resp. y_t , $t \in \mathcal{T}_2$) inférieurs à z . Il s'agit ainsi, sous l'hypothèse de stationnarité mentionnée précédemment, d'une estimation que le phénomène aléatoire sous-jacent prenne une valeur inférieure à z .

La stationnarité d'une série temporelle est une propriété stipulant que la distribution de la variable considérée ne change pas en fonction du temps. Sous une hypothèse de continuité supplémentaire sur la distribution du phénomène aléatoire sous-jacent, il est possible de comparer des estimations de densités à la place d'estimations de fonctions de répartition. Il s'agira par exemple de comparer des *estimations à noyaux* des densités sous-jacentes :

$$\hat{f}_x(z) = \frac{1}{h|\mathcal{T}_1|} \sum_{t \in \mathcal{T}_1} \mathsf{K}\left(\frac{z - x_t}{h}\right) \quad \text{et} \quad \hat{f}_y(z) = \frac{1}{h|\mathcal{T}_2|} \sum_{t \in \mathcal{T}_2} \mathsf{K}\left(\frac{z - y_t}{h}\right), \quad (4.2)$$

où $h \geq 0$ est la largeur de la *fenêtre* spécifiant le degré de lissage de la densité estimée et K est un *noyau*. Un choix usuel est le noyau Gaussien donné par :

$$\mathsf{K}(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad z \in \mathbb{R}.$$

Dans la pratique, les fonctions quantiles estimées \hat{F}_x^{-1} et \hat{F}_y^{-1} seront également utilisées. Pour une proportion p donnée, un quantile d'ordre p peut ainsi être calculé par $\hat{F}_x^{-1}(p) = q_{xp}$ pour la série $(x_t)_{t \in \mathcal{T}_1}$ (resp. $\hat{F}_y^{-1}(p) = q_{yp}$ pour la série $(y_t)_{t \in \mathcal{T}_2}$). À partir d'un vecteur de proportions (p_i) et grâce aux fonctions \hat{F}_x^{-1} et \hat{F}_y^{-1} , il est ainsi possible d'obtenir deux séries de quantiles correspondants aux séries $(x_t)_{t \in \mathcal{T}_1}$ et $(y_t)_{t \in \mathcal{T}_2}$. La représentation graphique par diagramme quantiles-quantiles permet de comparer graphiquement les quantiles de deux séries. L'égalité en distribution des deux séries temporelles correspond à la situation où les points du diagramme quantiles-quantiles sont, approximativement, sur la droite $y = x$.

Les outils décrits ci-dessus peuvent tous être implémentés à l'aide du langage de programmation R [65], lequel permet également aisément les représentations graphiques correspondantes, comme les diagrammes quantiles-quantiles ou les densités estimées données dans (4.2).

4.1.3 Correction d'une série temporelle à partir d'une autre

Si les étapes précédentes mettent en valeur une différence entre les distributions stationnaires des séries $(x_t)_{t \in \mathcal{T}_1}$ et $(y_t)_{t \in \mathcal{T}_2}$, il est possible de corriger une série à partir de l'autre en utilisant des versions empiriques de la *transformation en probabilité* et de la *transformation quantile*.

Soit X une variable aléatoire de fonction de répartition F_X , donnée par $F_X(z) = \Pr(X \leq z)$, $z \in \mathbb{R}$. Si F_X est continue, alors la *transformation en probabilité* nous indique que la variable aléatoire $F_X(X)$ suit une loi uniforme sur l'intervalle $]0, 1[$.

Soit F_X^{-1} la fonction quantile de X (qui est simplement l'inverse de F_X si F_X est continue strictement croissante) et soit U une variable aléatoire uniforme sur l'intervalle $]0, 1[$. La *transformation quantile* nous indique alors que la variable aléatoire $F_X^{-1}(U)$ a alors F_X pour fonction de répartition.

Des versions empiriques de ces transformations peuvent être utilisées par exemple pour corriger une série temporelle de simulations $(y_t)_{t \in \mathcal{T}_2}$ à partir d'observations *in situ* $(x_t)_{t \in \mathcal{T}_1}$. La première étape consiste à calculer les fonctions de répartition \hat{F}_x et \hat{F}_y données par (4.1). La version corrigée de $(y_t)_{t \in \mathcal{T}_2}$ est alors $(\tilde{y}_t)_{t \in \mathcal{T}_2}$, où

$$\tilde{y}_t = F_x^{-1}(F_y(y_t)), \quad t \in \mathcal{T}_2. \quad (4.3)$$

La série temporelle résultante $(\tilde{y}_t)_{t \in \mathcal{T}_2}$ aura alors approximativement la même distribution stationnaire que la série $(x_t)_{t \in \mathcal{T}_1}$.

4.2 Comparaisons des simulations avec des observations *in situ*

À l'aide des outils présentés dans la section précédente, nous pouvons évaluer l'adéquation entre les séries simulées disponibles et des observations *in situ* en certains lieux. Nous considérons ainsi successivement les points correspondants aux bouées d'Anglet et de Saint-Jean-de-Luz pour les séries d'états de mer ainsi que le point du marégraphe de Socoa pour les données de niveaux d'eau. Cette comparaison entraînera, en cas de détection de biais de simulation, la correction des séries temporelles issues des simulations afin de permettre une analyse des dommages plus pertinentes dans le dernier chapitre.

4.2.1 Au point de la bouée d'Anglet

Le point correspondant à la bouée d'Anglet est situé à environ cinq kilomètres de la côte, en face du phare de Biarritz. Les séries temporelles d'états de mer, appelées **BOUEE.ANG** et **WWMIL.ANG** d'une part, et **BOUEE.ANG** et **HOM.ANG** d'autre part, seront comparées dans cette section. Rappelons que **BOUEE.ANG** fait référence à des observations d'états de mer *in situ* au point de la bouée d'Anglet et que les séries temporelles **WWMIL.ANG** et **HOM.ANG** font référence aux simulations numériques décrites dans les sections 3.1.2.ii et 3.1.2.i, respectivement, restreintes au point de la bouée d'Anglet. Le point de la bouée d'Anglet étant situé à une profondeur de 50 mètres, cette valeur sera utilisée pour le calcul du flux P défini par la formule (2.6). Pour le Runup $R_{2\%}$ défini par (2.7), la valeur de la pente β a été prise égale à 0.007%. Cette valeur correspond en fait à la mesure de la pente de la Grande Plage de Biarritz obtenue entre l'hiver 2017 et l'hiver 2019 [15]. Bien qu'il soit peu probable que cette valeur soit constante au cours du temps, elle sera néanmoins utilisée pour les trois jeux de données considérés dans cette section.

Les comparaisons entre valeurs simulées et observées au point de la bouée d'Anglet ont été effectuées pour les variables H_s , T_p , T_{m10} , P et $R_{2\%}$ introduites dans la section 2.4. Le tableau 4.1 donne les statistiques descriptives pour chacune de ces variables et les trois jeux de données considérés. Le tableau 4.2 fournit les indicateurs présentés dans la section 4.1.1. Une inspection du tableau 4.1 suggère que les

CHAPITRE 4. VALIDATION ET CORRECTION DES SÉRIES TEMPORELLES
D'ÉTATS DE MER ET DE NIVEAUX D'EAU

| Var. | BOUEE.ANG | | | WWMII.ANG | | | HOM.ANG | | |
|------------|-----------|---------|--------|-----------|---------|--------|---------|---------|--------|
| | Min. | Max | Moy. | Min. | Max | Moy. | Min. | Max | Moy. |
| H_s | 0.13 | 8.9 | 1.732 | 0.097 | 8.754 | 1.546 | 0.148 | 9.264 | 1.609 |
| T_p | 2.11 | 21.90 | 10.54 | 2.807 | 20.994 | 10.788 | X | X | X |
| T_{m10} | 3.2 | 18.10 | 8.987 | 2.383 | 17.196 | 9.359 | 2.430 | 16.690 | 9.219 |
| P | 0.064 | 678.573 | 24.406 | 0.019 | 322.951 | 15.945 | 0.055 | 624.096 | 19.475 |
| $R_{2\%}$ | 0.092 | 4.806 | 1.136 | 0.091 | 4.938 | 1.206 | X | X | X |
| θ_m | 0.0 | 360.0 | 305.7 | 226.4 | 344.8 | 303.2 | 0.0 | 360.0 | 304.0 |

TABLEAU 4.1 – Résumés statistiques des séries temporelles d'états de mer disponibles au point de la bouée d'Anglet.

| Simu. | WWMII.ANG | | | | HOM.ANG | | | |
|------------|-----------|-----------|--------|-------|---------|-----------|--------|-------|
| | Biais | Biais (%) | RMSE | R^2 | Biais | Biais (%) | RMSE | R^2 |
| H_s | 0.112 | 6.524 | 0.484 | 0.730 | 0.075 | 4.625 | 0.302 | 0.846 |
| T_{m10} | -0.314 | -3.512 | 3.027 | 0.014 | -0.465 | -5.282 | 1.077 | 0.788 |
| P | 7.494 | 31.019 | 27.737 | 0.370 | 2.789 | 14.106 | 13.743 | 0.648 |
| θ_m | 2.678 | 0.877 | 19.064 | 0.094 | 2.032 | 0.664 | 17.408 | 0.271 |

TABLEAU 4.2 – Biais de simulation absolu et en pourcentage, racine de l'erreur quadratique moyenne (RMSE) et coefficient de détermination (R^2) pour les simulations WWMII et Homere au point de la bouée d'Anglet pour les variables H_s , T_{m10} et P.

ordres de grandeurs pour les variables H_s , T_p , T_{m10} et $R_{2\%}$ sont globalement comparables d'un jeu de données à l'autre, ce qui semble indiquer que les plages des valeurs simulées sont relativement réalistes. En ce qui concerne le flux P, on constate néanmoins qu'il est clairement sous-estimé, en moyenne, dans la simulation WWMII. Ce dernier point est confirmé par le tableau 4.2 qui montre un biais de simulation très important pour le flux P, en particulier pour la simulation WWMII qui a ainsi tendance à très fortement sous-estimer le flux en moyenne.

Les panneaux de gauche des figures 4.1, 4.2, 4.3 et 4.4 montrent les diagrammes quantiles-quantiles obtenus en comparant les valeurs simulées et observées de H_s , T_{m10} , P et $R_{2\%}$, respectivement. Ces diagrammes ont été réalisés comme décrit dans la section 4.1.2. Les panneaux de droite de ces mêmes figures donnent les densités estimées des valeurs observées et des valeurs issues des deux simulations.

Pour chacune des variables H_s , T_{m10} , P et $R_{2\%}$, les densités estimées à partir d'observations et de simulations semblent relativement proches. Ce sont les diagrammes quantiles-quantiles qui sont les plus informatifs et révèlent globalement une tendance des simulations à sous-estimer les quantiles modérés à élevés des quatre variables considérées. En inspectant la figure 4.1, on observe clairement que H_s est sous-estimée pour des hauteurs significatives de vagues supérieures à 2.5 m. Sans surprise au vue de la formule (2.6), la sous-estimation du flux P est encore plus marquée dans la figure 4.3 car H_s intervient au carré dans le calcul de P. La simulation WWMII est particulièrement affectée avec des quantiles simulés approximativement 50% inférieurs aux quantiles observés lorsque $P > 200 \text{ J} \cdot \text{m}^{-1} \cdot \text{s}^{-1}$. La sous-estimation du flux P par les simulations est également illustrée dans la figure 4.3 où

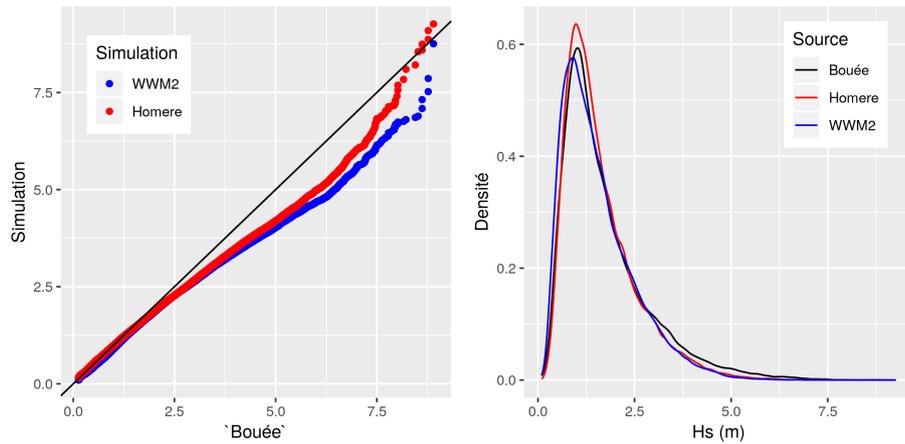


FIGURE 4.1 – Diagramme quantiles-quantiles pour H_s (en m) obtenu en comparant les données d'observation et les données de simulation au point de la bouée d'Anglet (à gauche) et densités estimées des valeurs observées et des valeurs issues des deux simulations (à droite).

un zoom a été effectué sur les densités estimées du flux entre $300 \text{ J} \cdot \text{m}^{-1} \cdot \text{s}^{-1}$. Par exemple, pour la simulation WWMII, la densité estimée devient nulle à partir de $340 \text{ J} \cdot \text{m}^{-1} \cdot \text{s}^{-1}$, alors que la densité estimée des valeurs observées apparaît comme clairement non nulle au moins jusqu'à $P = 400 \text{ J} \cdot \text{m}^{-1} \cdot \text{s}^{-1}$. En ce qui concerne la période T_{m10} , on constate à partir de la figure 4.2 que l'adéquation entre valeurs observées et simulées est bien meilleure, même si les périodes faibles sont globalement légèrement sur-estimées et les périodes élevées sont globalement légèrement sous-estimées. Enfin, à partir de la figure 4.4, on constate la relative bonne adéquation entre quantiles observés et simulés du Runup $R_{2\%}$ pour la simulation WWMII. Rappelons que le Runup $R_{2\%}$ n'a pas pu être calculé pour la simulation Homere car cette dernière ne contient pas la valeur de la période pic T_p .

En conclusion, au point de la bouée d'Anglet, même si certaines variables semblent être correctement simulées, la plupart des variables sont sous-estimées, et parfois, très fortement. Les biais de simulation, correspondant aux grandes vagues notamment, nous incitent à tenter d'utiliser la méthode de correction fondée sur les transformations en probabilité et quantile présentée dans la section 4.1.3. Plusieurs essais de telles corrections seront décrits dans la section 4.3.

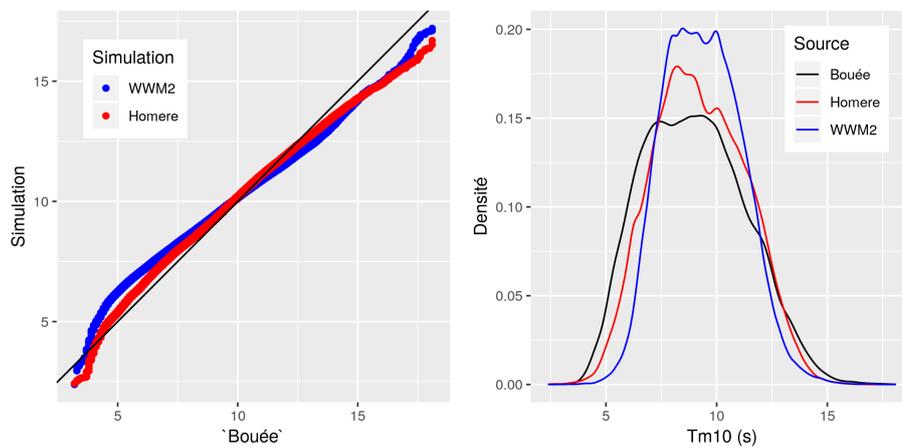


FIGURE 4.2 – Diagramme quantiles-quantiles pour T_{m10} (en s) obtenu en comparant les données d'observation et les données de simulation au point de la bouée d'Anglet (à gauche) et densités estimées des valeurs observées et des valeurs issues des deux simulations (à droite).

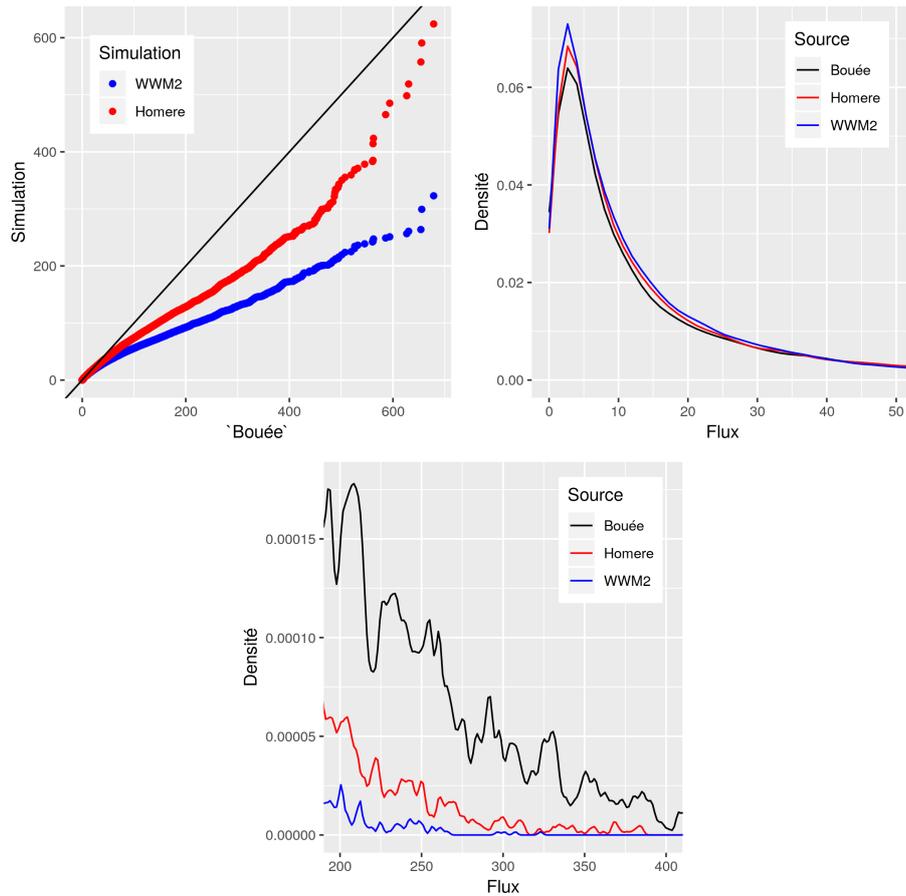


FIGURE 4.3 – Diagramme quantiles-quantiles pour P (en $J.m^{-1}.s^{-1}$) obtenu en comparant les données d'observation et les données de simulation au point de la bouée d'Anglet (à gauche) et densités estimées des valeurs observées et des valeurs issues des deux simulations (à droite). La figure du bas donne les densités estimées entre 300 et 400 $J.m^{-1}.s^{-1}$.

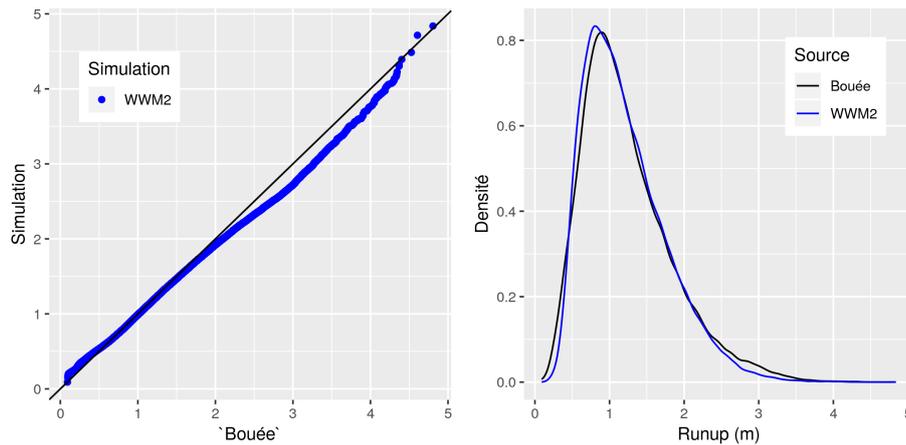


FIGURE 4.4 – Diagramme quantiles-quantiles pour $R_{2\%}$ (en m) obtenu en comparant les données d'observation et les données de simulation au point de la bouée d'Anglet (à gauche) et densités estimées des valeurs observées et des valeurs issues des deux simulations (à droite).

4.2.2 Au point de la bouée de Saint-Jean-de-Luz

La bouée de Saint-Jean-de-Luz est située face au port de la ville, quasiment en face de la digue de l'Artha et du fort de Socoa. Elle se situe à environ 2 kilomètres du littoral et est ancrée à une profondeur de 20 mètres. Nous rappelons que la simulation Homere ayant un très faible recouvrement temporel avec les données de la bouée, les données issues de cette simulation ne pourront pas être exploitées ici. Ainsi, seules les séries temporelles de la bouée et de la simulation WWMII seront comparées : il s'agit des jeux de données **BOUEE.SJL** et **WWMII.SJL** présentés dans la section 3.1. Étant donnée l'absence d'information sur les pentes de plages de la baie de Saint-Jean-de-Luz, nous n'avons accès à aucune valeur de pente de plage β , ce qui implique qu'aucun calcul de Runup ne sera envisagé dans cette section.

Pour le point de la bouée de Saint-Jean-de-Luz, les séries temporelles observées et simulées par WWMII ont été comparées pour les variables H_s , T_{m10} , T_p et le flux P . Dans le tableau 4.3, les statistiques descriptives de chacune des variables sont données pour **BOUEE.SJL** et **WWMII.SJL**. Dans le tableau 4.4, on retrouve les indicateurs présentés dans la section 4.1.1. De même que pour la bouée d'Anglet dans la section précédente, nous pouvons constater, à partir du tableau 4.3, que les ordres de grandeurs des variables H_s , T_{m10} et T_p sont plutôt comparables entre les observations et la simulation WWMII. En revanche, le flux P semble encore une fois largement sous-estimé. Cela est confirmé par l'inspection de la valeur du biais de simulation donné dans le tableau 4.4. Contrairement à la situation observée au point de la bouée d'Anglet, la variable H_s apparaît comme légèrement sur-estimée, en moyenne, par la simulation. Un tel résultat pourrait s'expliquer par la faible longueur des séries temporelles d'observations pour le point de la bouée de Saint-Jean-de-Luz (la bouée a été mise en service en avril 2013).

Les parties gauches des figures 4.5, 4.6 et 4.7 montrent les diagrammes quantiles-quantiles obtenus en comparant les valeurs d'observation et de simulation par WWMII pour H_s , T_{m10} et le flux P . Les parties droites de ces figures correspondent aux densités estimées des valeurs observées et des valeurs de la simulation WWMII. Ces figures ont été réalisées comme expliqué dans la section 4.1.2.

En terme de diagrammes quantiles-quantiles, les résultats obtenus pour le point de la bouée de Saint-Jean-de-Luz sont proches de ceux observés pour le point de la bouée d'Anglet. En effet, les vagues au dessus de 2.5 m semblent globalement sous-estimées par WWMII (même si, en moyenne, WWMII semble sur-estimer légèrement les hauteurs de vagues significatives). Pour la période T_{m10} , les périodes très faibles ainsi que les périodes supérieures à 10 s sont sous-estimées par le modèle. Entre ces deux valeurs, le modèle semble sur-estimer légèrement la période T_{m10} . Comme pour la bouée d'Anglet, la simulation WWMII sous-estime très fortement le flux P .

Les conclusions pour le point de la bouée de Saint-Jean-de-Luz sont ainsi très proches de celles pour le point de la bouée d'Anglet. Malgré des ordres de grandeur corrects pour les variables, on constate une sous-estimation marquée des grandes vagues. Ces biais de simulation vont nous inciter à utiliser la méthode à base des

| Variable | BOUEE.SJL | | | WWMII.SJL | | |
|-----------|-----------|---------|--------|-----------|---------|--------|
| | Min. | Max | Moy. | Min. | Max | Moy. |
| H_s | 0.140 | 8.560 | 1.611 | 0.097 | 9.246 | 1.601 |
| T_{m10} | 3.500 | 19.500 | 9.389 | 2.383 | 17.208 | 9.450 |
| T_p | 2.100 | 38.900 | 10.850 | 2.808 | 20.786 | 10.842 |
| P | 0.103 | 545.164 | 22.972 | 0.021 | 345.258 | 16.393 |

TABLEAU 4.3 – Résumés statistiques des séries temporelles d'états de mer disponibles au point de la bouée de Saint-Jean-de-Luz.

| Simu. | WWMII.SJL | | | |
|-----------|-----------|--------|-----------|-------|
| | Variable | Biais | Biais (%) | RMSE |
| H_s | -0.108 | -6.461 | 0.447 | 0.838 |
| T_{m10} | 0.163 | 1.758 | 3.443 | 0.067 |
| T_p | -0.310 | -2.899 | 1.867 | 0.645 |
| P | 5.917 | 23.312 | 23.463 | 0.477 |

TABLEAU 4.4 – Biais de simulation absolu et en pourcentage, racine de l'erreur quadratique moyenne (RMSE) et coefficient de détermination (R^2) pour les simulations WWMII et Ho-mere au point de la bouée de Saint-Jean-de-Luz pour les variables H_s , T_{m10} , T_p et P.

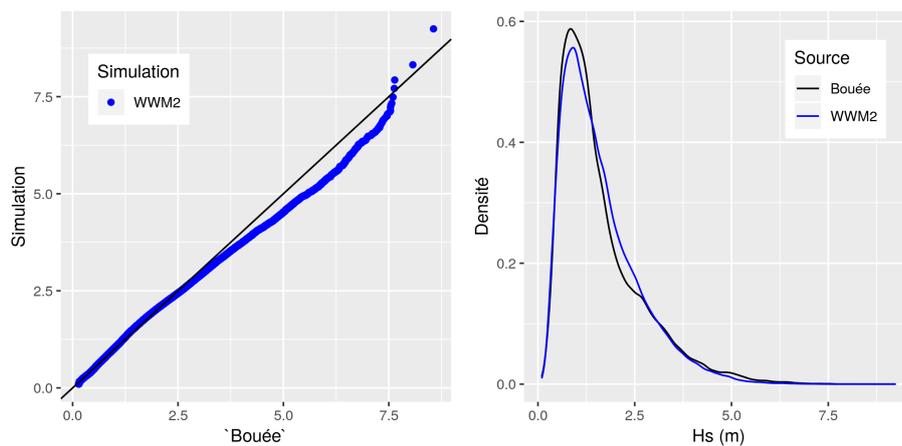


FIGURE 4.5 – Diagramme quantiles-quantiles pour H_s (en m) obtenu en comparant les données d'observation et les données de simulation au point de la bouée de Saint-Jean-de-Luz (à gauche) et densités estimées des valeurs observées et des valeurs issues des deux simulations (à droite).

transformations en probabilité et quantile vues dans la section 4.1.3 pour corriger les séries temporelles simulées. Les essais de correction seront décrits dans la section 4.3.

4.2.3 Au point du marégraphe de Socoa

Pour l'étude des données de niveaux d'eau, seules les données issues du marégraphe de Socoa ont été utilisées. Les séries temporelles **C1.SOCOA** et **C2.SOCOA**

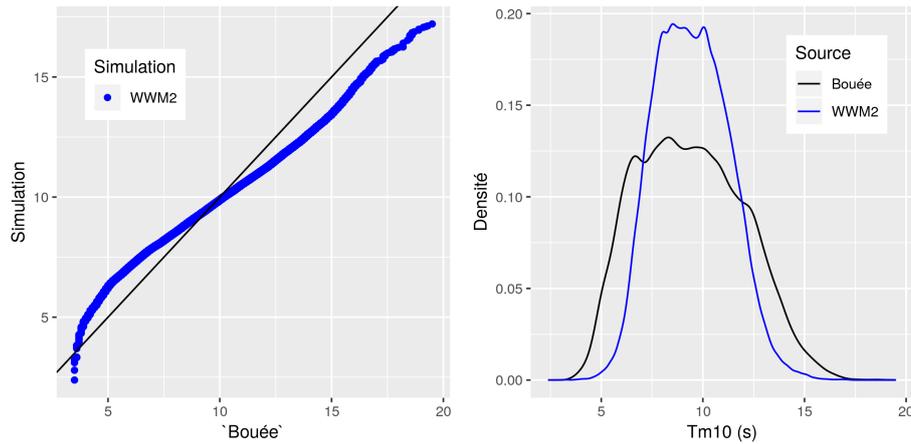


FIGURE 4.6 – Diagramme quantiles-quantiles pour T_{m10} (en s) obtenu en comparant les données d'observation et les données de simulation au point de la bouée de Saint-Jean-de-Luz (à gauche) et densités estimées des valeurs observées et des valeurs issues des deux simulations (à droite).

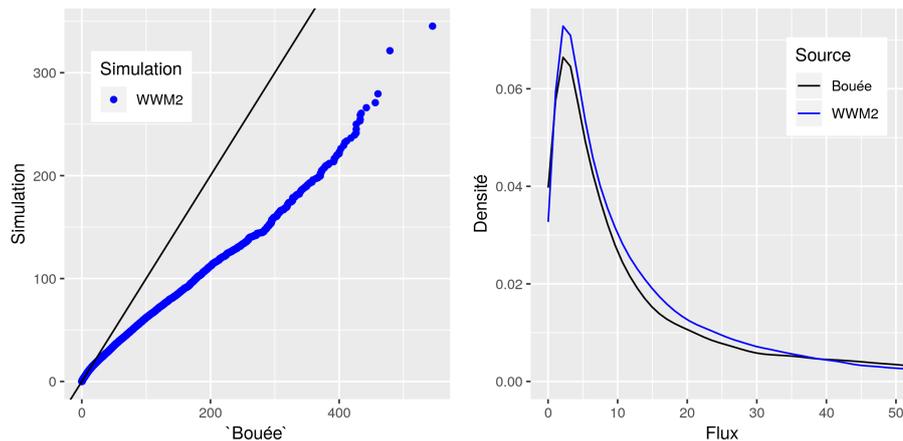


FIGURE 4.7 – Diagramme quantiles-quantiles pour le flux P (en $J.m^{-1}.s^{-1}$) obtenu en comparant les données d'observation et les données de simulation au point de la bouée de Saint-Jean-de-Luz (à gauche) et densités estimées des valeurs observées et des valeurs issues des deux simulations (à droite).

de niveaux d'eau η calculées dans la section 3.2.2 ont été ainsi comparées à la série temporelle du marégraphe **MAR.SOCOA** présentée dans la section 3.2.1.

Les indicateurs vus à la section 4.1.1 ont été calculés pour comparer les données **MAR.SOCOA** mesurées par le marégraphe avec les jeux de données de réanalyse **C1.SOCOA** et **C2.SOCOA**. Les résultats sont consignés dans le tableau 4.5. Les indicateurs indiquent une très bonne proximité à la réalité pour les deux méthodes de calcul avec un léger avantage pour **C1.SOCOA** qui est calculé avec la formule (3.1).

À l'aide de la méthode de la section 4.1.2, les fonctions de densités ont été estimées pour les trois jeux de données et représentées sur la figure 4.8. De plus, un diagramme quantiles-quantiles a été réalisé. Sur les deux représentations, la proximité entre les distributions des réanalyses et du marégraphe est évidente, avec, encore

| Stats | C1.SOCOA | C2.SOCOA |
|----------------|----------|----------|
| Biais | 0.0 | 0.0 |
| Biais (%) | 0.0 | 0.019 |
| RMSE | 0.079 | 0.116 |
| R ² | 0.993 | 0.985 |

TABLEAU 4.5 – Biais de simulation absolu et en pourcentage, racine de l'erreur quadratique moyenne (RMSE) et coefficient de détermination (R²) pour les simulations WWMII et Homere au point du marégraphe de Socoa pour la variable η (en m).

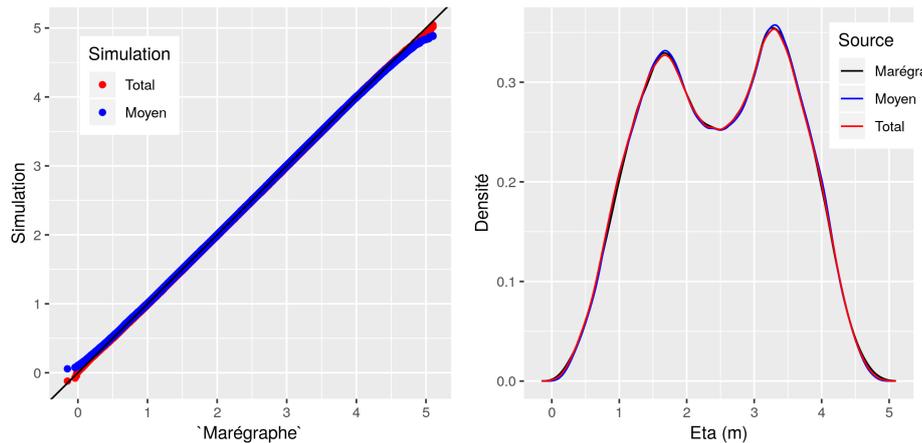


FIGURE 4.8 – Diagrammes quantiles-quantiles (à gauche) et fonctions de densités estimées (à droite) pour η (en m) pour les données au point du marégraphe de Socoa.

une fois, un léger avantage pour **C1.SOCOA**.

Les données de réanalyses de niveaux d'eau établies à l'aide de la méthode présentée dans la section 3.2.2 montrent une grande proximité aux données du marégraphe. La série temporelle calculée **C1.SOCOA** apparaît comme la meilleure représentation de la réalité. La comparaison de la série temporelle du marégraphe avec **C2.SOCOA** a également permis de vérifier que la méthode vue dans la section 3.2.2, qui se traduit par l'utilisation de la formule (3.1), donne effectivement de meilleurs résultats. Ainsi, pour la variable η , le jeu de données **C1.SOCOA** pourra être utilisé tel quel pour l'analyse des dommages qui sera effectuée dans le chapitre 6.

4.3 Séries temporelles corrigées en distribution

La série temporelle de niveaux d'eau **C1.SOCOA** est une suffisamment bonne représentation de la réalité pour être utilisée dans le cadre d'une étude de longue durée sur les événements de submersion. Concernant les données représentant les états de mer, nous avons vu dans les sections 4.2.1 et 4.2.2 que les simulations ne semblaient pas approcher suffisamment bien la réalité, ayant notamment tendance à sous-estimer les grandes vagues. La simulation WWMII couvre une fenêtre temporelle suffisamment longue pour convenir à l'étude faite dans ce document, ce

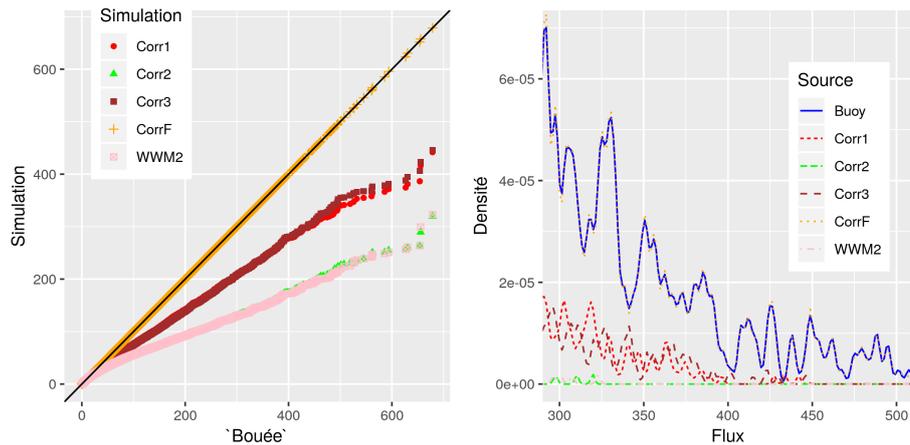


FIGURE 4.9 – Diagrammes quantiles-quantiles (à gauche) et fonctions de densités (à droite) pour le flux P (en $J.m^{-1}.s^{-1}$) pour la série temporelle de la bouée, de la simulation WWMII et les quatre corrections testées au point de la bouée d'Anglet.

qui n'est malheureusement pas le cas de la simulation Homere. Ainsi, pour pouvoir exploiter la simulation WWMII, une série de tests de correction, basée sur les transformations en probabilité et quantile a été effectuée selon la méthode présentée dans la section 4.1.3 et résumée par l'équation (4.3). Les résultats seront présentés ci-après en détail pour le point de la bouée d'Anglet.

Pour tester l'efficacité d'une correction en distribution fondée sur l'équation (4.3), les méthodes de comparaisons quantiles-quantiles et les graphes des densités estimées vus dans la section 4.1.2 seront encore une fois utilisés. Le flux P semblant d'un point physique une variable cruciale pour potentiellement expliquer les dégâts liés à la submersion, les résultats des corrections seront considérés comme satisfaisants quand une série temporelle de la variable du flux P aura une distribution similaire à celle du flux calculé à partir des données observées. Plusieurs essais de correction sur les variables impliquées dans le calcul du flux (2.6), H_s et T_{m10} , ont été effectués pour le point de la bouée d'Anglet :

- **Corr1** : correction en distribution de H_s uniquement;
- **Corr2** : correction en distribution de T_{m10} uniquement;
- **Corr3** : correction en distribution de H_s et de T_{m10} .

La figure 4.9 montre dans sa partie gauche les diagrammes quantiles-quantiles obtenus en comparant les valeurs observées du flux P au point de la bouée d'Anglet avec les valeurs simulées par WWMII, ainsi qu'avec les valeurs des méthodes de correction ci-dessus. Pour résumer, la série temporelle du flux de **BOUEE.ANG** est comparée successivement à celle issue de **WWMII.ANG**, de **Corr1**, de **Corr2** et enfin de **Corr3**. La partie droite de la figure correspond aux densités estimées à partir des mêmes séries temporelles du flux.

Comme on peut le constater sur la figure 4.9, la correction sur T_{m10} seulement (**Corr2**, en vert sur les graphes) n'améliore presque pas la distribution des valeurs du flux. En ce qui concerne la correction sur les hauteurs de vagues uniquement (**Corr1**, en rouge sur les graphes), elle conduit à une amélioration sensible du flux

tout en restant une sous-estimation du flux réel. La conclusion est très similaire pour **Corr3** (en marron sur les graphes).

Au vu de ces résultats, nous avons décidé de procéder à des corrections en distribution de cinq variables pour la série temporelle issue de la simulation WWMII : H_s , T_{m10} , θ_m , P et $R_{2\%}$. Sans surprise, la distribution corrigée du flux **CorrF** (en orange sur la figure) coïncide alors parfaitement sur la figure 4.9 avec la distribution des valeurs observées au point de la bouée d'Anglet. Un exemple de comparaison entre les données mesurées par la bouée (en rouge), la simulation (en bleu) et la simulation corrigée (en vert) est visible sur la figure 4.10. Pour cet exemple, la bouée a un pas de temps de 30 minutes alors que les données simulées ont un pas de temps de 3 heures. La correction de H_s montre, pour une grande partie des dates de cet exemple, une amélioration de la représentation de cette variable. Néanmoins, nous pouvons noter que le fait de corriger la distribution peut diminuer légèrement la précision pour certains points. Les événements de tempête peuvent être engendrés par différents types de dépression océanique, pour cette correction en distribution aucune différence significative de qualité de correction n'a pu être observée sur l'intégralité du jeu de données.

Une procédure similaire a été effectuée pour les données simulées par WWMII au point de la bouée de Saint-Jean-de-Luz. Les résultats, équivalents, sont résumés par la figure 4.11.

Finalement, dans la suite du document, les jeux de données **WWMII.ANG** et **WWMII.SJL** correspondront ainsi aux versions de ces séries temporelles où chacune des variables H_s , T_{m10} , T_p , P , θ_m et $R_{2\%}$ a été corrigée par transformation en probabilité et quantile à l'aide de l'équation (4.3). Pour former les jeux de données finaux de variables d'aléa, la série temporelle **C1.SOCOA** sera utilisée conjointement et synchronisée à ces deux jeux de données. Notons que ces données couvrent la période 1949-2015.

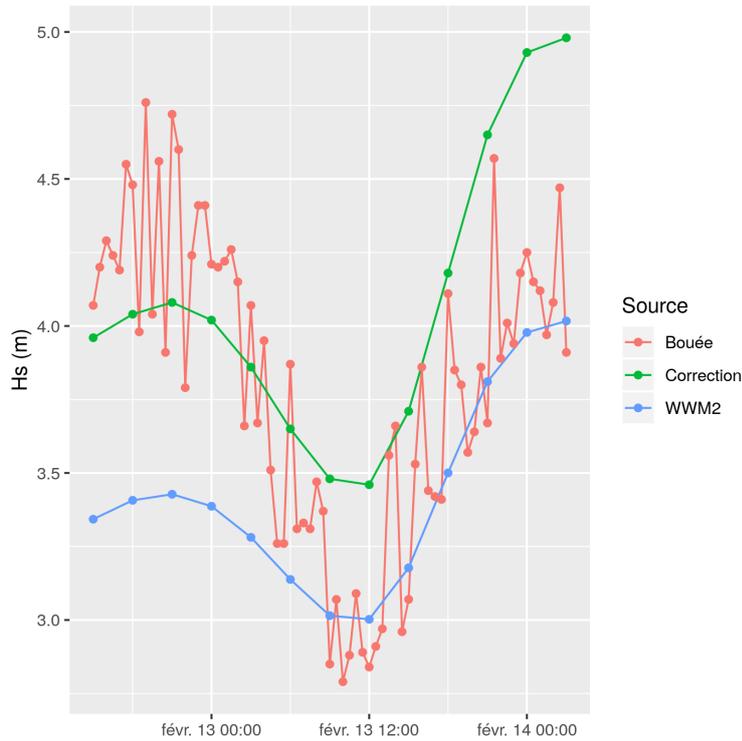


FIGURE 4.10 – Valeurs de H_s en m pour la bouée (en rouge), **WWMILANG** avant (en bleu) et après (en vert) correction au niveau du point de la bouée d'Anglet entre le 12 et le 14 février 2011.

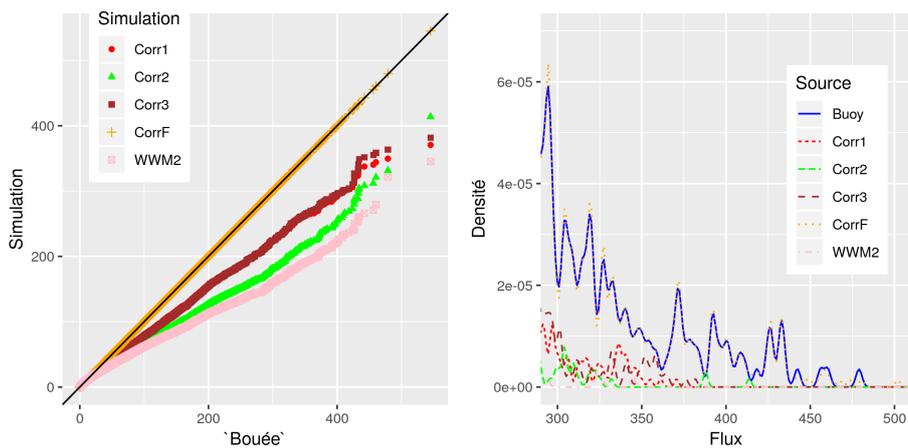


FIGURE 4.11 – Diagrammes quantiles-quantiles (à gauche) et fonctions de densités (à droite) pour le flux P (en $J.m^{-1}.s^{-1}$) pour la série temporelle de la bouée, de la simulation WWMII et les quatre corrections testées au point de la bouée de Saint-Jean-de-Luz.