

FIGURE 10 – Comparaison des valeurs réelles et prédites du logarithme de la composante sud, pour le ballon 2

2.4.2 Ajout de retards dans les modèles

L'ajout de retards de la variable cible parmi les variables explicatives est sans doute le moyen le plus simple et le plus direct de prendre en compte la dépendance temporelle des observations. On s'attend en effet à ce que la valeur de la variable cible dépende des valeurs précédentes. Le diagramme d'autocorrélation (ACF) de la série, semblable d'un ballon à un autre, (voir Figure 11) suggère d'ajouter quatre retards.

Toutefois, en ce qui concerne le modèle linéaire, l'ajout de variable retardées invalide l'hypo-

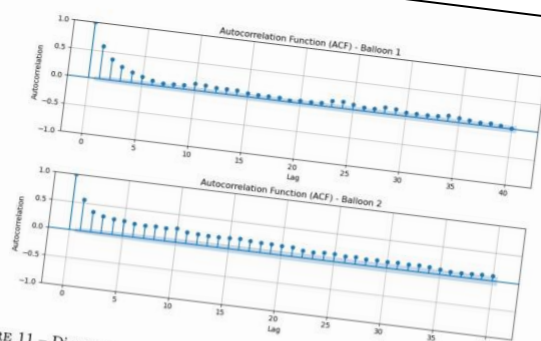


FIGURE 11 – Diagramme ACF de la série *gvmf_v_south* pour les deux premiers ballons

thèse d'indépendance des observations et provoque en général un biais dans l'estimation des coefficients; il sera donc utilisé seulement à titre indicatif. Pour ce modèle, comme pour les autres (RandomForest, XGBoost), on s'attend à avoir de meilleures prédictions.

Effectivement, la Figure 12 montre qu'avec l'ajout de quatre retards, les scores RMSE s'améliorent, et que la hiérarchie des modèles ne semble pas modifiée. Mais surtout, on peut voir sur les Figures 14 et 15 que les pics sont bien mieux prédits par tous les modèles, y compris le modèle linéaire.

Toutefois, la Figure 13 montre qu'une importance considérable est accordée aux variables retardées, au détriment des autres variables explicatives. Autrement dit, la prédiction s'améliore, mais au détriment du but recherché, à savoir la mise en évidence de relations entre les variables explicatives et le phénomène climatique que l'on cherche à modéliser : les ondes de gravité.