

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

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 La ajout de resaus de la variable cible parim les variables expineatives est sans doute le moyen le plus simple et le plus direct de prendre en compte la dépendance temporelle des observations. ne plus simple et le plus direct de prendre en compte la dependance temporelle des observations.

On s'attend en effet à ce que la valeur de la variable cible dépende des valeurs précédentes. Le On s'atteno en emet a ce que la valeur de la variable ciule depende des valeurs precedences. Le diagramme d'autocorrélation (ACF) de la série, semblable d'un ballon à un autre, (voir Figure 11) suggère d'ajouter quatre retards.

ggere a ajouter quatre retaras. Toutefois, en ce qui concerne le modèle linéaire, l'ajout de variable retardées invalide l'hy-

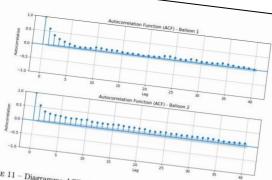


FIGURE 11 – Diagramme ACF de la série gvmf_v_south pour les deux premiers ballons

pothèse d'indépendance des observations et provoque en général un biais dans l'estimation des pourese à maependance des observations et provoque en general un mais dans i estimation des coefficients; il sera donc utilisé seulement à titre indicatif. Pour ce modèle, comme pour les autres coencients; il sera donc utilise semement a ette minicatal. Four te moute (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'amé-Entectivement, la rigure 12 montre qu'avec i ajout de quatre retards, les scores rande s'anne-liorent, et que la hiérarchie des modèles ne semble pas modifiée. Mais surtout, on peut voir sur les norent, et que la merarcine des modeles ne semble pas modifiec. 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

reaire.

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, targees, au detriment des autres variables expireatives. Autrement dit, la premetion s'amenore, 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.