

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{]t; \infty[} (T_x)$$

ressources-actuarielles.net



Chaire DAMI

27/03/2019

Construction de tendances de mortalité avec un réseau de neurones RVFL

Version 1.2

Corentin GUILBAUD
Frédéric PLANCHET

- La construction de projections d'espérances de vie fait l'objet de nombreux travaux depuis l'article fondateur de Lee et Carter (LEE et CARTER [1992]).
- Dans la perspective d'extrapoler dans le futur les tendances observées dans le passé, la majorité des approches proposées se basent sur une « surface de mortalité », mesurant les forces de mortalité par âge et année du moment, qu'il s'agit donc d'extrapoler dans la dimension temporelle.
- Les modèles inspirés par Lee et Carter commencent par réduire la dimension en réalisant un ACP puis en extrapolant une ou deux séries temporelles associées à la projection sur les axes principaux.
- Bongaarts (BONGAARTS [2004]) a proposé une démarche différente basée sur des ajustements paramétriques par année du moment et de l'extrapolation des coefficients estimés chaque année.

- Dans BONGAARTS [2004], l'auteur utilise toutefois une représentation paramétrique assez frustre (modèle de Thatcher) qui ne permet pas d'englober tous les âges. De plus, il limite son extrapolation à 2 paramètres sur 3 en les traitant de manière indépendante, ce qui est une approximation discutable.
- On part ici d'une représentation paramétrique plus riche, inspirée par les travaux de DEBONNEUIL [2015] conduisant à un modèle à 6 paramètres (7 dont 1 figé).
- Les 6 séries temporelles obtenues étant fortement corrélées, on propose ici d'utiliser un réseaux de neurones monocouche particulier, le RVFL (*Random Vector Functional Link*), pour prendre en compte efficacement cette dépendance, en s'inspirant de l'application faite par COUSIN et al. [2018] pour les modèles de taux d'intérêt.
- Le travail complet est présenté dans GUILBAUD [2018].

Ajustements paramétriques

- Le modèle retenu est le suivant

$$\mu(x) = Old(x) + Mid(x) + Young(x).$$

avec les composantes

$$Young(x) = e^{a_{young} + b_{young}x},$$

$$Mid(x) = \beta e^{-\frac{(x-m)^2}{\sigma}},$$

$$Old(x) = \mu_{asympt} \frac{1}{1 + e^{-ax-b}}.$$

- L'ajustement peut être effectué par maximum de vraisemblance avec une vraisemblance binomiale ou minimisation d'une erreur quadratique. L'EMV a été retenu, étant plus stable.
- La mise en œuvre numérique est délicate et doit pouvoir être répétée plusieurs dizaines de fois ; Il n'existe pas de solution explicite évidente pour maximiser la log-vraisemblance. L'expression du gradient est complexe.

Ajustements paramétriques

- Nous avons donc eu recours aux algorithmes implémentés dans la bibliothèque TensorFlow de Abadi et al. [2015] développée par Google afin d'implémenter un algorithme de descente de gradient.
- L'algorithme Adam a été retenu (KINGMA et BA [2014]). Il garde en mémoire la trace du gradient et de son carré à l'aide de deux vecteurs m et v mis à jour à chaque pas de temps :

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla_{\theta} f(\theta_t)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) (\nabla_{\theta} f(\theta_t))^2$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t$$

- L'ajustement est effectué sur la plage 1955-2015 à partir des données de la base HMD.

Extrapolation par réseau de neurones

- BONGAARTS [2004] a utilisé des modèles ARIMA pour projeter dans le temps les deux paramètres de son modèle. Avec son approche l'espérance de vie des femmes devient significativement plus faible que celle des hommes au bout d'une vingtaine d'année.
- Les modèles ARIMA ne reflètent pas les dépendances entre les paramètres et on se tourne donc vers les réseaux RVFL, qui permettent de prendre en compte cette dépendance de manière assez directe.
- On considère le modèle RVFL dans lequel la réponse Y , de dimension n , est expliquée par p prédicteurs $Z^{(j)}$ (chacun de dimension n) de la manière suivante :

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j Z_i^{(j)} + \sum_{l=1}^L \gamma_l g \left(\sum_{j=1}^p W^{(j,l)} Z_i^{(j)} \right) + \epsilon_i$$

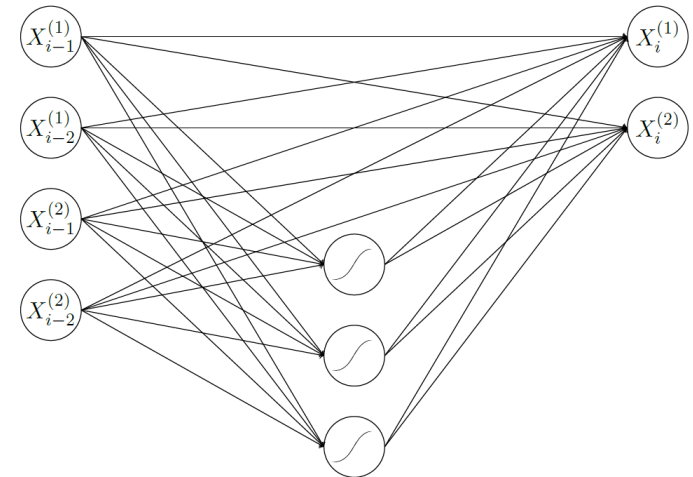
- On projette les coefficients différenciés à l'ordre un pour rendre les séries stationnaires.

Extrapolation par réseau de neurones

- g est la fonction d'activation, L , le nombre de nœuds et $W^{(j,i)}$ des poids ; cette spécification fait donc intervenir une composante linéaire et une composante non linéaire basée sur les prédicteurs transformés

$$\Phi(\mathbf{Z})^{(l)} = g \left(\sum_{j=1}^p W^{(j,l)} Z_i^{(j)} \right)$$

- Les poids $W^{(j,i)}$ ne sont pas estimés mais choisis avec un générateur de nombres uniformes quasi-aléatoire (Sobol), l'idée étant qu'ils soient « bien uniformément répartis ».
- Les paramètres à estimer sont donc les β_j et γ_l , ainsi que le nombre de nœuds L .



Couche d'entrée

Couche cachée

Couche de sortie

FIGURE 3.5 – Schéma d'illustration d'un réseau RVFL avec $p = 2$, $k = 2$ et $L = 3$.

Extrapolation par réseau de neurones

- Le principe est de faire une régression linéaires sur les prédicteurs (initiaux et transformés), donc d'utiliser un critère de moindres carrés ; toutefois, pour éviter les problèmes de conditionnement liés à la colinéarité des variables explicatives (d'autant plus forte avec une série temporelle), on impose une pénalisation Ridge :

$$\sum_{j=1}^p \beta_j^2 \leq u \qquad \sum_{l=1}^L \gamma_l^2 \leq v$$

- La résolution des conditions du premier ordre conduit à une expression explicite des paramètres à partir de l'inversion de matrices, pour les coefficients (λ_1, λ_2) issus de u et v dans le Lagrangien donnés.
- Ce cadre général est adapté pour les séries temporelles de la manière suivante.

Extrapolation par réseau de neurones

- On dispose de p séries temporelles observées à n dates, on souhaite construire une prédiction de ces p séries à h dates $n+1, \dots, n+h$. On s'appuie sur un historique de $k < n$ valeurs de la série.

- La matrice des prédicteurs est alors constituée $n-k$ lignes de p bloc de k valeurs retardées de la forme

$$\left(X_{n-i}^{(j_0)}, \dots, X_{k+1-i}^{(j_0)} \right)^T$$

- La matrice des prédicteurs transformés est de taille $(n-k) \times (k \times p + L)$

- Avec $p=2$ séries observées à $n=5$ dates, $k=2$ valeurs retardées et $L=3$ nœuds on a :

$$Y = \begin{pmatrix} X_{t_5}^{(1)} & X_{t_5}^{(2)} \\ X_{t_4}^{(1)} & X_{t_4}^{(2)} \\ X_{t_3}^{(1)} & X_{t_3}^{(2)} \end{pmatrix} \quad X = \begin{pmatrix} X_{t_4}^{(1)} & X_{t_3}^{(1)} & X_{t_4}^{(2)} & X_{t_3}^{(2)} \\ X_{t_3}^{(1)} & X_{t_2}^{(1)} & X_{t_3}^{(2)} & X_{t_2}^{(2)} \\ X_{t_2}^{(1)} & X_{t_1}^{(1)} & X_{t_2}^{(2)} & X_{t_1}^{(2)} \end{pmatrix} \quad W = \begin{pmatrix} W^{(1,1)} & W^{(1,2)} & W^{(1,3)} \\ W^{(2,1)} & W^{(2,2)} & W^{(2,3)} \\ W^{(3,1)} & W^{(3,2)} & W^{(3,3)} \\ W^{(4,1)} & W^{(4,2)} & W^{(4,3)} \end{pmatrix}$$

Application numérique

- Le choix des hyperparamètres ne peut ici être effectué comme dans COUSIN et al. [2018] avec des fenêtres glissantes du fait d'un nombre de points insuffisant dans chaque série (seulement 61 points).
- Le protocole utilisé pour chaque modèle est alors le suivant :
 - Les modèles sont calibrés à l'aide des points obtenus strictement avant 2006.
 - Les points entre 2006 et 2015 servent de test afin de choisir les hyperparamètres. Ce choix est basé sur une recherche par grille : nous entraînons r fois les modèles sur l'intervalle [1955;2005] pour r jeux d'hyperparamètres différents, nous retenons le jeu ayant l'erreur quadratique la plus faible sur l'intervalle [2006; 2015].
 - Nous sommes obligés de prendre un échantillon de test relativement petit. En effet, une bosse sur la mortalité autour de 50 ans est apparue au début des années 2000. Si nous l'excluons de l'ensemble d'entraînement, le modèle ne la fera pas apparaître. Dès lors si la prédictions du modèle ne parait pas assez stable, nous nous réservons le droit d'augmenter la régularisation des modèles pour lisser les projections. Une fois les hyperparamètres choisis, nous entraînons le modèle sur toutes les données pour projeter la mortalité au delà de 2015.

Application numérique

- Le choix des hyperparamètres ne peut ici être effectué comme dans COUSIN et al. [2018] avec des fenêtres glissantes du fait d'un nombre de points insuffisant dans chaque série (seulement 61 points).
- Le protocole utilisé pour chaque modèle est alors le suivant :
 - Les modèles sont calibrés à l'aide des points obtenus strictement avant 2006.
 - Les points entre 2006 et 2015 servent de test afin de choisir les hyperparamètres. Ce choix est basé sur une recherche par grille : nous entraînons r fois les modèles sur l'intervalle [1955;2005] pour r jeux d'hyperparamètres différents, nous retenons le jeu ayant l'erreur quadratique la plus faible sur l'intervalle [2006; 2015].
 - Nous sommes obligés de prendre un échantillon de test relativement petit. En effet, une bosse sur la mortalité autour de 50 ans est apparue au début des années 2000. Si nous l'excluons de l'ensemble d'entraînement, le modèle ne la fera pas apparaître. Dès lors si la prédiction du modèle ne paraît pas assez stable, nous nous réservons le droit d'augmenter la régularisation des modèles pour lisser les projections. Une fois les hyperparamètres choisis, nous entraînons le modèle sur toutes les données pour projeter la mortalité au delà de 2015.

Application numérique

- On trouve un ajustement optimal avec $L=20$ neurones, conduisant aux résultats suivants (table unisexe) :

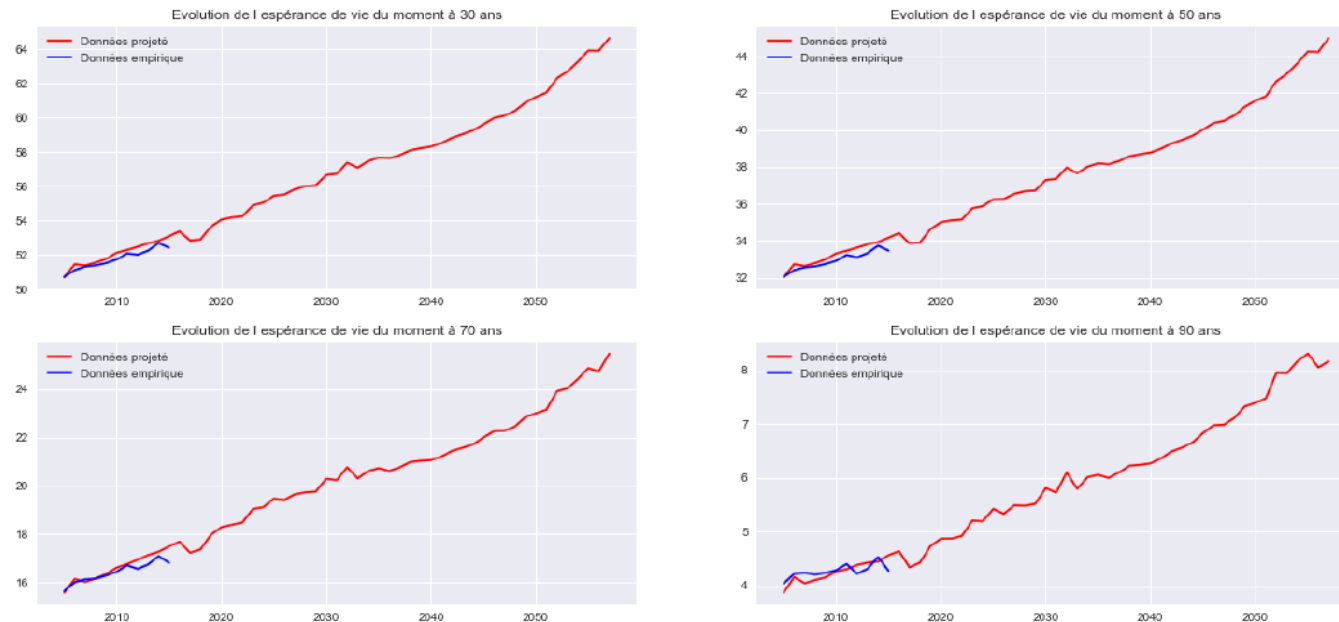


FIGURE 3.11 – Espérances de vie du moment empirique (en bleu) et projetée par RVFL(1) (en rouge) sur les séries différenciées à 30, 50, 70 et 90 ans.

- L'analyse des résidus montre toutefois un problème de surapprentissage après 2005 ; pour contourner cette difficulté, on cherche à réduire la dimension.

Application numérique

- On fait une ACP sur les séries différenciées en conservant les 3 premiers facteurs, qui expliquent 80 % de la variance.



FIGURE 3.18 – Séries initiales (en bleu) et reconstruites (en rouge) par ACP.

Application numérique

- Toutefois, l'ajustement ainsi obtenu induit un biais de sous-estimation des espérances de vie du moment

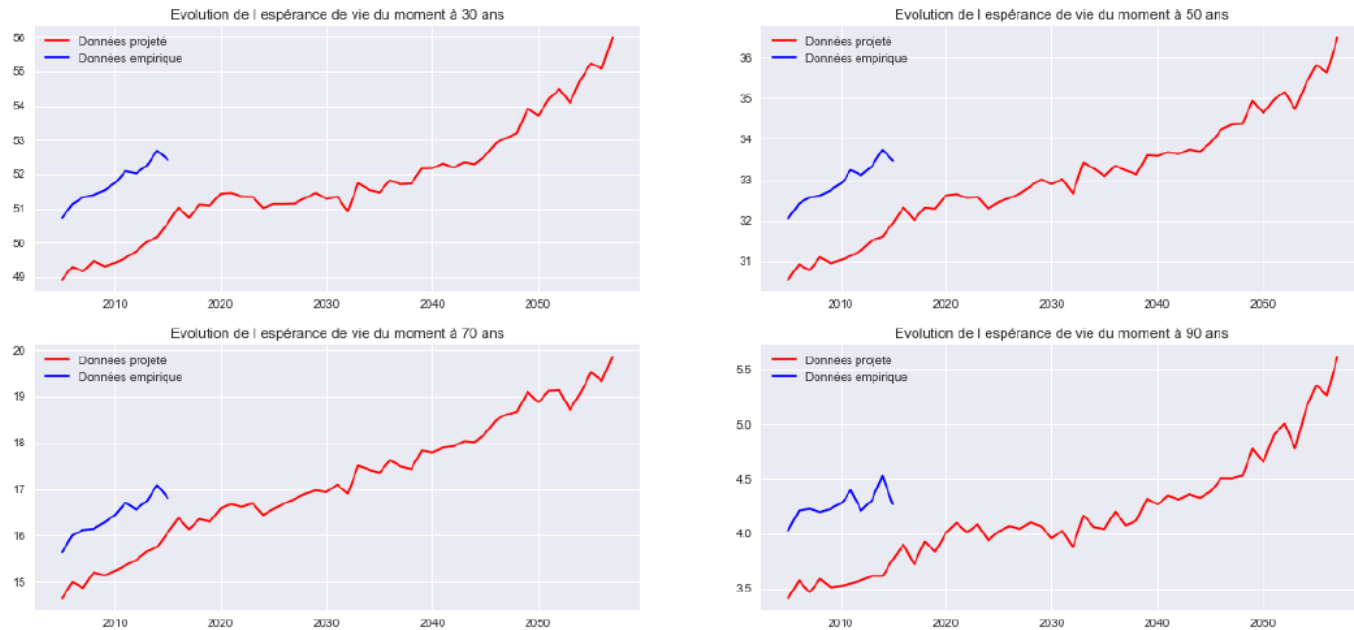
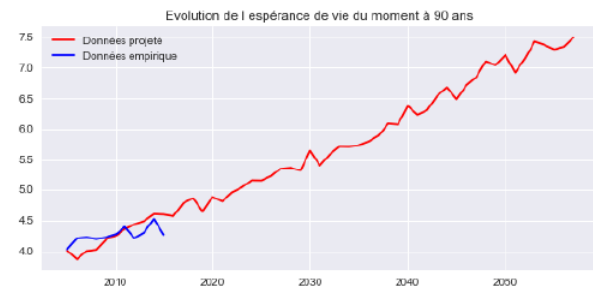
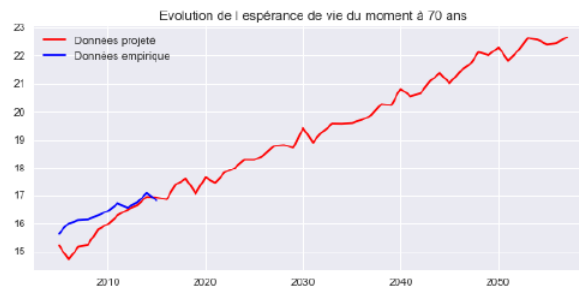
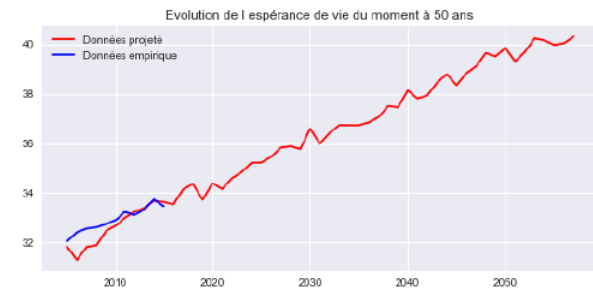
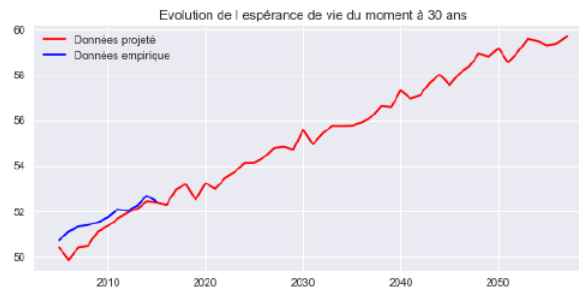


FIGURE 3.20 – Espérances de vie du moment empirique et projetée par RVFL(3) sur les séries différenciées avant ACP à 30, 50, 70 et 90 ans.

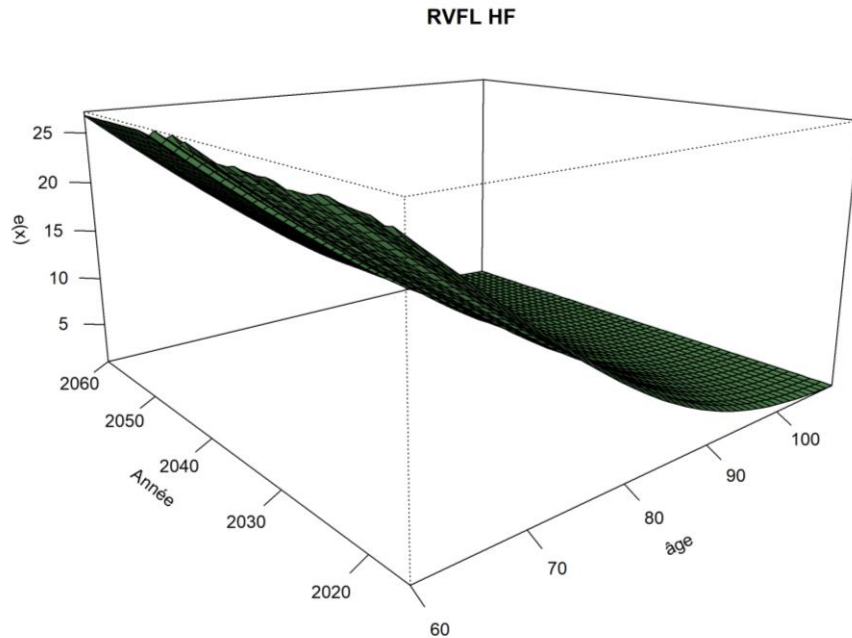
Application numérique

- Il est possible de corriger cette erreur en inversant l'ordre du traitement des données, c'est à dire en commençant par appliquer l'ACP aux séries brutes, puis en différenciant les composantes retenues.
- Nous ne conservons que les deux premières composantes principales qui expliquent 79% de la variance. Avec $L = 40$ nœuds les taux de mortalité sont projetés en continuité avec les données observées.

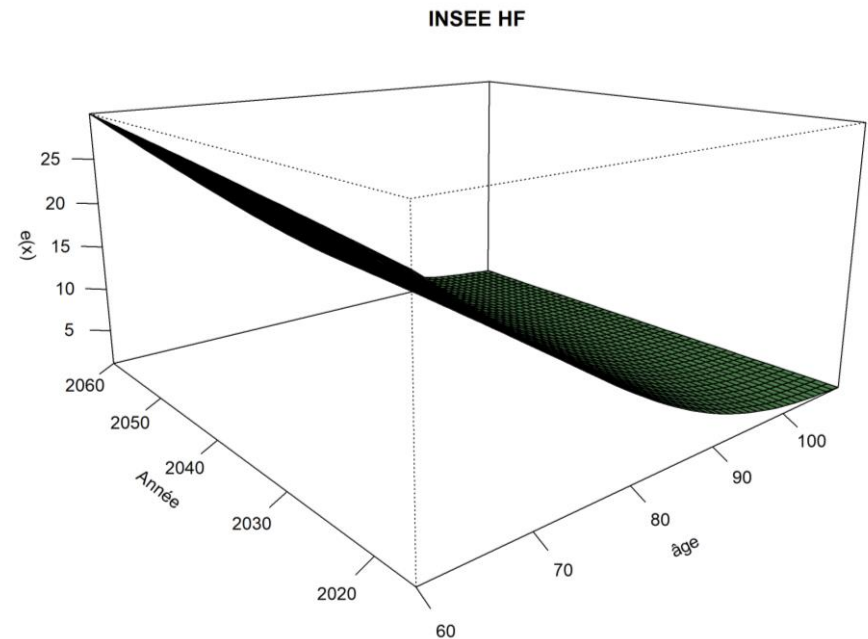


Application numérique

■ L'allure des prédictions obtenues est la suivante

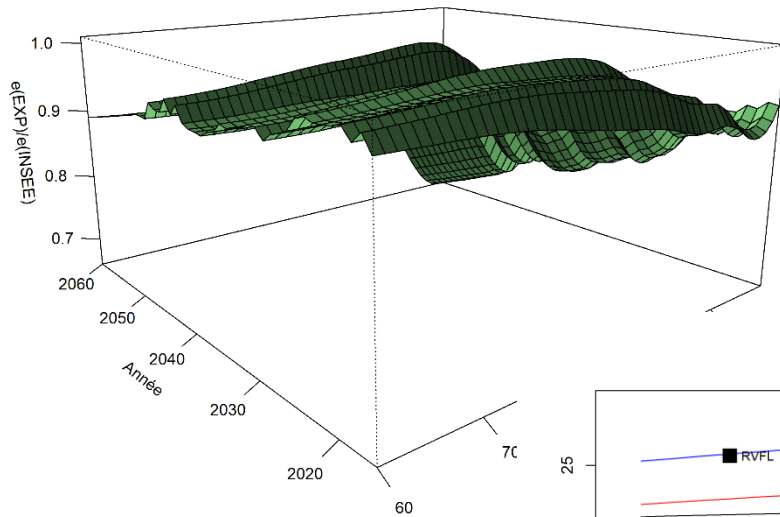


La régularité de la surface est moins forte que, par exemple, les **tables INSEE 2060**



Application numérique

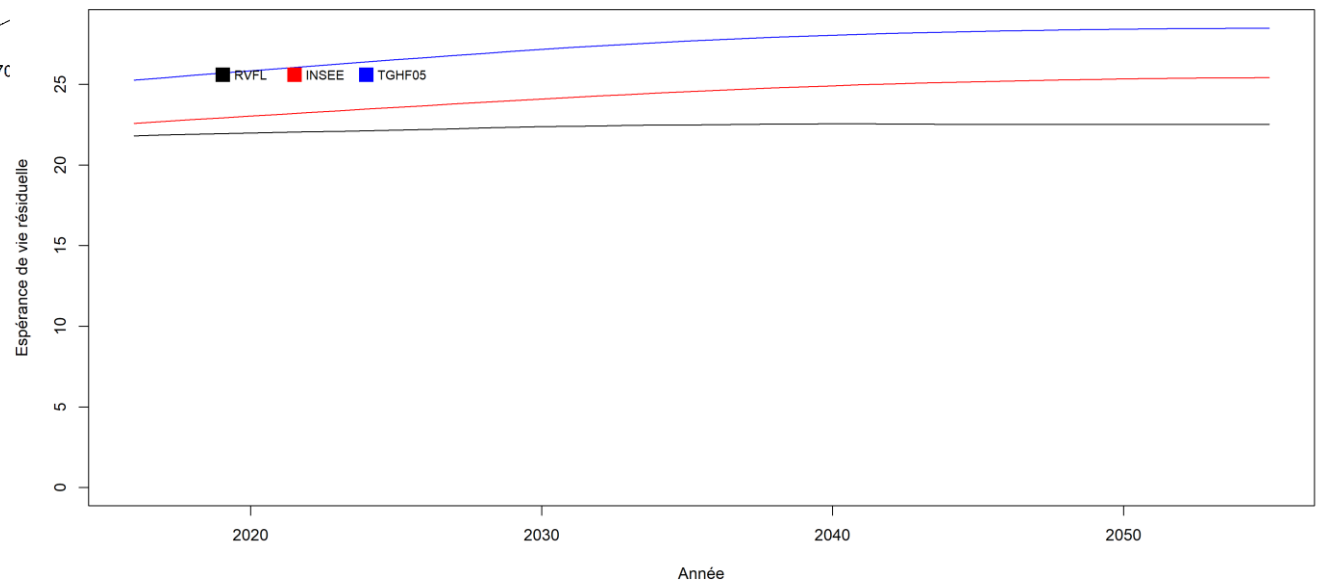
■ ■ La comparaison des espérances de survie prospectives avec celle de l'INSEE conduit à :



On obtient donc des prédictions inférieures à celles issues des tables INSEE 2060, avec un écart ayant tendance à croître avec le temps.

L'explication de cet écart reste à construire.

Comparaison des espérances de vie résiduelles à 65 ans (unisexe)



Conclusion

- Les modèles de type RVFL fournissent un outil performant pour réaliser des projections de séries temporelles multivariées avec une forte dépendance entre les composantes.
- Toutefois, du fait du petit nombre de points dans chaque série, le calibrage est délicat, des ajustements sont nécessaires pour construire des prédictions cohérentes.
- En particulier, on doit, comme dans LC, utiliser une ACP pour réduire la dimension, ce que l'on cherchait à éviter initialement.
- Enfin, l'absence de cadre probabiliste empêche de fournir simplement des intervalles de confiance de prédiction (ce qui est par exemple simple dans un modèle de type log-Poisson)
- La suite des travaux va se focaliser sur l'extrapolation directe des espérances de vie du moment, *a priori* plus simples à projeter que des forces de mortalité.

- BONGAARTS J. [2004] “Long-range trends in adults mortality : Models and projection methods”, Population Council .
- COUSIN A., MOUDIKI T., PLANCHET F. [2018] Multiple time series forecasting using quasi-randomized functional link neural networks, *Risks*, 2018, 6(1), 22; doi:10.3390/risks6010022.
- DEBONNEUIL E. [2015] Modèle paramétrique de mortalité en fonction de l'âge, pour des applications à des portefeuilles de retraite, Mémoire d'actuaire, ISFA.
- GUILBAUD C. [2018] Nouveaux modèles d'analyse et de projection de la mortalité, application a la population française, Mémoire d'actuaire, Dauphine.
- KINGMA D. , BA J. [2014], “Adam : A method for stochastic optimization”, Conference paper at ICLR 2015.
- LEE R. D., CARTER L. [1992] “Modeling and forecasting us mortality”, *Journal of the American statistical association* 87(419), 659{671.