Modélisation des séries temporelles

Séance 3

Compléments sur ARIMA, processus SARIMA

Frédéric Sur

https://members.loria.fr/FSur/enseignement/modseries/

Séance 3

Modélisation des chroniques par (S)ARIMA

- soit la chronique est stationnaire
  (propriétés statistiques invariantes au cours du temps)
  → modélisation AR / MA / ARMA.
- soit elle ne l'est pas
  (tendance stochastique / marche aléatoire, ou tendance déterministe)
  → on commence par dériver pour stationnariser.

Méthode de Box-Jenkins : rappel

1. **Transformation** (éventuelle) de la chronique (généralement log) pour « se ramener » au modèle additif.
2. **Identification** des paramètres p, d, q.
   → identification des ordres p et q avec fonctions `Acf` et `Pacf` (fonction `tsdisplay`), éventuellement différentiée à l'ordre d au préalable.
3. **Estimation** des θ_j, ϕ_i, μ (ou `constant`) et σ.
   (fonction `Arima`)
4. **Validation** du modèle.
   → gaussianité des résidus (ACF, Portmanteau, graphe des résidus), significativité des coefficients, σ.
   (fonctions `tsdiag`, `coeftest`...)
5. **Prévision** du futur.
   (fonction `forecast`, aujourd'hui)

Prévision avec modèle ARIMA(p,d,q) (1)

Rappel : processus ARIMA

$$\Phi(B)\left((1-B)^d X_t - \mu\right) = \Theta(B)\varepsilon_t$$

ou :

$$\Psi(B)X_t = \theta_0 + \Theta(B)\varepsilon_t$$

c'est-à-dire *(pour simplifier, on suppose $\theta_0 = 0$)* :

$$X_t = \psi_1 X_{t-1} + \ldots + \psi_{p+d} X_{t-p-d} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

Prévision : on connaît (X_t) jusque la date $t = T$, on cherche une prévision $\widehat{X}_T(h)$ de X à un *horizon* de h après la date T.

Prévision avec modèle ARIMA(p,d,q) (2)

Pour tout t', on définit $\widehat{X}_{t'}(h) = \mathbb{E}(X_{t'+h} \,|\, (X_t)_{t \leqslant t'})$.
$\rightarrow$ projection : « meilleure » approximation de $X_{t'+h}$ par combinaison linéaire des $(X_t)_{t\leqslant t'}$ (cf. seance 2)

Remarques : pour tout $t' \leq T$

- $\widehat{X}_{t'}(h) = X_{t'+h}$ si $h \leqslant 0$.
- $\widehat{\varepsilon}_{t'}(h) = \mathbb{E}(\varepsilon_{t'+h} \,|\, (X_t)_{t\leqslant t'}) = \varepsilon_{t'+h}$ si $h \leqslant 0$ et $\widehat{\varepsilon}_{t'}(1) = 0$ car ε b.b.g.

De plus : $X_{t'+1} = \sum_{i=1}^{p+d} \psi_i X_{t'+1-i} + \varepsilon_{t'+1} - \sum_{j=1}^{q} \theta_j \varepsilon_{t'+1-j}$

Donc : $\widehat{X}_{t'}(1) = X_{t'+1} - \varepsilon_{t'+1}$ (par projection)

soit : pour tout t, $\varepsilon_t = X_t - \widehat{X}_{t-1}(1)$
$\rightarrow$ cela va permettre l'estimation des ε_t pour $t \leqslant T$

Mais comment calcule-t-on les prévisions $\widehat{X}_t(1)$, $\widehat{X}_t(2)$, etc ?

Prévision avec modèle ARIMA(p,d,q) (3)

$$X_{T+h} = \sum_{i=1}^{p+d} \psi_i X_{T+h-i} + \varepsilon_{T+h} - \sum_{j=1}^{q} \theta_j \varepsilon_{T+h-j}$$

Donc, avec $h \geqslant 0$:

$$\widehat{X}_T(h) = \sum_{i=1}^{p+d} \psi_i \widehat{X}_T(h-i) - \sum_{j=h}^{q} \theta_j \varepsilon_{T+h-j}$$

car $\mathbb{E}(\varepsilon_{T+h-j} | (X_t)_{t\leqslant T}) = \varepsilon_{T+h-j}$ si $j \geqslant h$ et $= 0$ sinon.
($\varepsilon_{T+h-j} \in \mathrm{Vect}(X_t)_{t\leqslant T}$ si $j \geqslant h$ et (ε_t) non corrélés).

Conséquence : formules d'actualisation avec les ψ_i, θ_j, ε_t :

$$\widehat{X}_T(1) = \sum_{i=1}^{p+d} \psi_i X_{T+1-i} - \sum_{j=1}^{q} \theta_j \varepsilon_{T+1-j}$$

$$\widehat{X}_T(2) = \psi_1 \widehat{X}_T(1) + \sum_{i=2}^{p+d} \psi_i X_{T+2-i} - \sum_{j=2}^{q} \theta_j \varepsilon_{T+2-j}$$

etc.

`forecast` **R** : I.C. pour $\widehat{X}_T(h)$ sous hypothèse de normalité des ε_t

Séance 3

Cas des chroniques périodiques, période π

Remarque 1 : X_t peut ne pas être stationnaire à cause d'un comportement du type :

$$X_t = S_t + u_t \qquad (S_t \text{ déterministe } \pi\text{-périodique})$$

ou

$$X_t = \mu + X_{t-\pi} + u_t \qquad (\text{cf marche aléatoire})$$

$\rightarrow$ on peut stationnariser en étudiant $(1 - B^\pi)X_t$.

Remarque 2 : des corrélations saisonnières (période π) peuvent être présentes dans la chronique X_t
(corrélations / corrélations partielles aux décalages de π, 2π, $3\pi\ldots$)

$\rightarrow$ ARMA *saisonnier* :

$$\phi(B^\pi)X_t = \theta(B^\pi)\varepsilon_t$$

Les processus SARIMA

Définition **: processus SARIMA$(p, d, q)(P, D, Q)_\pi$**

Ce sont les processus (X_t) :

$$\Phi_p(B)\Phi_P(B^\pi)\left((1-B)^d(1-B^\pi)^D X_t - \mu\right) = \Theta_q(B)\Theta_Q(B^\pi)\varepsilon_t$$

où $p, d, q, P, D, Q \geqslant 0$, π est la période de la saisonnalité et (ε_t) est un bruit blanc gaussien.

Autre notation : $(1-B)^d(1-B^\pi)^D X_t = \mu + \dfrac{\Theta_q(B)\Theta_Q(B^\pi)}{\Phi_p(B)\Phi_P(B^\pi)}\varepsilon_t$

Intérêt : traiter les chroniques non-stationnaires, avec tendance et saisonnalité ou comportement style « marche aléatoire ».

Remarque : SARIMA = ARIMA particulier, mais la factorisation limite le nombre de coefficients à estimer.
(cf *parcimonie*, *rasoir d'Ockham*)

Exemples

1. processus SARIMA$(1, 0, 2)(1, 1, 0)_4$:
(chronique trimestrielle, période annuelle)

$$(1-\phi_1 B)(1-\phi'_1 B^4)(1-B^4)X_t = \theta_0 + (1-\theta_1 B - \theta_2 B^2)\varepsilon_t$$

ou

$$(1-\phi_1 B)(1-\phi'_1 B^4)\left((1-B^4)X_t - \mu\right) = (1-\theta_1 B - \theta_2 B^2)\varepsilon_t$$

2. processus SARIMA$(0, 1, 1)(1, 0, 0)_{12}$:
(chronique mensuelle, période annuelle)

$$(1-\phi'_1 B^{12})(1-B)X_t = \theta_0 + (1-\theta_1 B)\varepsilon_t$$

ou

$$(1-\phi'_1 B^{12})\left((1-B)X_t - \mu\right) = (1-\theta_1 B)\varepsilon_t$$

Identification des modèles

TP précédent : identification des ordres des processus ARIMA selon ACF et PACF.

TP aujourd'hui : identification des ordres des processus SARIMA : on **regarde ACF et PACF aux décalages π, 2π, $3\pi\ldots$**
(en pratique, pour saisonnalité annuelle : décalages 12 et 24)

Important : on cherche des modèles *simples*...

Cas AR/MA + SAR/SMA

Remarque : on commence par regarder ACF/PACF pour « petits décalages » ($h \leqslant 6$), on identifie AR/MA, puis ACF/PACF pour $h = 12, 24, 36$ et identification SAR/SMA.
(pour se débarrasser de l'influence des corrélations « court termes » sur la composante saisonnière)

Exemple : $X_t = (1 - \theta_1 B)(1 - \theta'_1 B^{12})\varepsilon_t$
alors : $X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta'_1 \varepsilon_{t-12} + \theta_1 \theta'_1 \varepsilon_{t-13}$

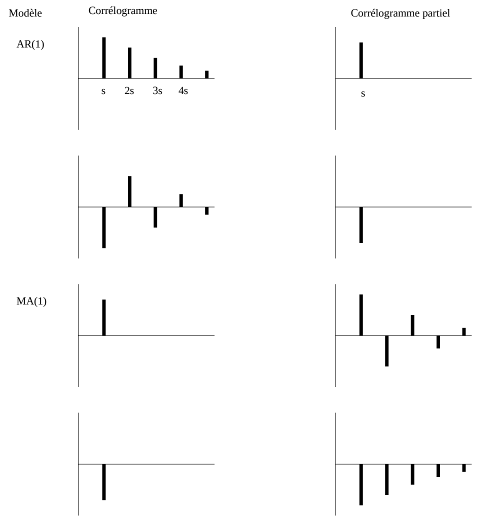
$$\text{donc : } \begin{cases} \gamma(11) = \text{Cov}(X_t, X_{t-11}) = \theta_1 \theta'_1 \\ \gamma(12) = \text{Cov}(X_t, X_{t-12}) = -\theta'_1 - \theta_1^2 \theta'_1 \\ \gamma(13) = \text{Cov}(X_t, X_{t-13}) = \theta_1 \theta'_1 \end{cases}$$

Mais si on identifie un premier modèle AR de résidus Y :
$X_t = (1 - \theta_1 B) Y_t$, donc $Y_t = (1 - \theta'_1 B^{12})\varepsilon_t$

$$\text{et : } \begin{cases} \gamma(11) = \text{Cov}(Y_t, Y_{t-11}) = 0 \\ \gamma(12) = \text{Cov}(Y_t, Y_{t-12}) = -\theta'_1 \\ \gamma(13) = \text{Cov}(Y_t, Y_{t-13}) = 0 \end{cases}$$

$\rightarrow$ ACF plus lisible pour identifier SMA

ACF et PACF pour (S)AR(1) et (S)MA(1)



Séance 3

1. Méthode de Box-Jenkins
   - Rappels
   - Prévision
2. Processus SARIMA
3. Quelques « règles » et compléments
   - Retour sur la significativité des paramètres
   - Retour sur la transformation par passage au log
   - Tests de Dickey-Fuller
   - Astuces et critères pour l'identification des modèles
4. Conclusion

Retour sur la significativité des paramètres du modèle

On estime un paramètre du modèle à partir de la chronique, réalisation d'un processus aléatoire :
$\rightarrow$ variable aléatoire $\overline{X}$ d'espérance μ
(μ= « vraie » valeur si estimateur non biaisé)
$\rightarrow$ l'estimation par `Arima` donne **une** valeur x (une réalisation)

L'erreur-type (*standard error* `se`) est l'estimation de l'écart-type de l'estimateur (calculé par `Arima`)

Exemple : estimation de la moyenne μ d'un échantillon i.i.d.
$\overline{X} = \frac{1}{N} \sum_{k=1}^{N} X_k$
si σ écart-type des X_k alors $\text{std}(\overline{X}) = \sigma / \sqrt{N}$

on calcule x et `se` empiriques (estimations de μ et $\text{std}(\overline{X})$)

z-test

Si ε_t est un b.b.g., et $\overline{X}$ est l'estimateur d'un paramètre du modèle (par exemple selon maximum de vraisemblance), la statistique de test

$$Z = (\overline{X} - \mu)/\mathrm{std}(\overline{X})$$

est distribuée selon une loi normale centrée réduite.

Test d'hypothèse (z-test) : $\mathcal{H}_0 : \mu = 0$
on calcule la p-valeur : $\Pr_{\mathcal{H}_0}(|Z| > |x|/\mathrm{se})$
Z distribuée selon une loi de Student (ou normale, cf. `help coeftest`)

Si p-valeur $< 5\%$, alors on rejette $\mathcal{H}_0$
$\rightarrow$ *il est peu probable que les fluctuations d'échantillonnage seules puissent expliquer une valeur aussi grande de $|x|/se$,*
$\mu \neq 0$ est sans doute une meilleure explication
$\rightarrow$ le paramètre est dit *significatif*
(c'est-à-dire significativement non nul)

Intervalles de confiance

Risque α $(= 5\%)$
$Z = (\overline{X} - \mu)/\mathrm{std}(\overline{X})$ suit une loi normale

Soit $z_\alpha > 0$ t.q. $\Pr(|Z| < z_\alpha) = 1 - \alpha$
(exemple : si $\alpha = 5\%$, alors $z_\alpha \simeq 2$)

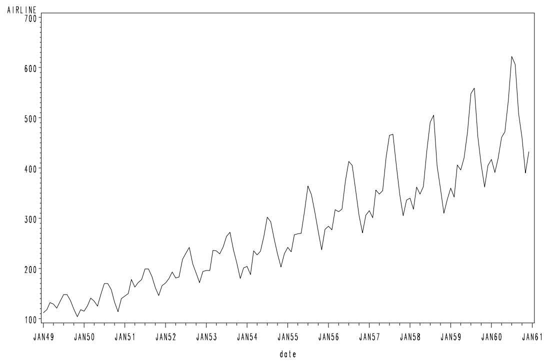
alors : $\Pr(-z_\alpha \leqslant (\overline{X} - \mu)/\mathrm{std}(\overline{X}) \leqslant z_\alpha) = 1 - \alpha$
d'où : $\Pr(\overline{X} - z_\alpha \mathrm{std}(\overline{X}) \leqslant \mu \leqslant \overline{X} + z_\alpha \mathrm{std}(\overline{X})) = 1 - \alpha$
$\rightarrow$ il y a donc une probabilité $1 - \alpha$ que μ soit dans l'intervalle (aléatoire...) $\overline{X} \pm z_\alpha \mathrm{std}(\overline{X})$

Conclusion : l'intervalle de confiance du paramètre estimé au risque α est : $x \pm z_\alpha \mathrm{se}$
$\rightarrow$ fonction `coefci`

Remarque : 0 est dans l'I.C. $\Leftrightarrow$ le paramètre est non-significatif
Preuve : 0 dans I.C. $\Leftrightarrow$ $x - z_\alpha \mathrm{se} < 0$ et $x + z_\alpha \mathrm{se} > 0$
$\Leftrightarrow$ $|x|/\mathrm{se} < z_\alpha$
$\Leftrightarrow$ $\Pr_{\mathcal{H}_0}(Z < |x|/\mathrm{se}) < 1 - \alpha$
$\Leftrightarrow$ $\Pr_{\mathcal{H}_0}(Z > |x|/\mathrm{se}) > \alpha$

Remarque : transformation de (X_t)...

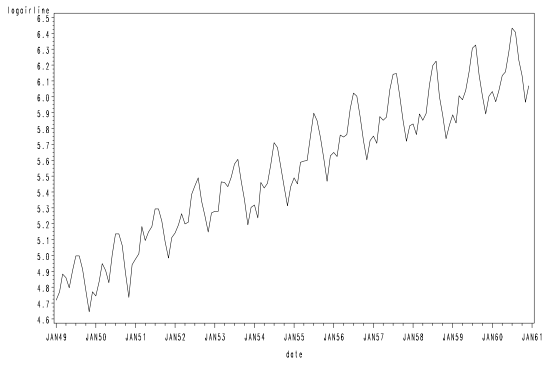
Exemple : chronique `airline`



$\rightarrow$ passage au log...

Passage au log

Exemple : logarithme de la chronique `airline`



$\rightarrow$ modèle additif avec tendance.

Question : *quid* de la prévision sur `airline` ?

Étude de $Y_t = \log(X_t)$

Hypothèse : $Y_\tau \sim \mathcal{N}(\widehat{Y}_\tau, \widehat{\sigma}_\tau^2) \quad (\tau > T)$

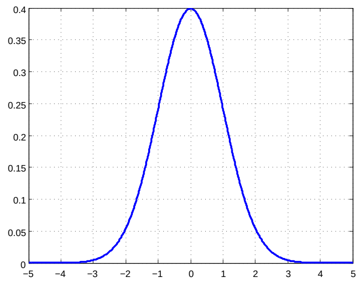
`forecast` : fournit la prévision $\widehat{Y}_\tau$
$+$ intervalle de confiance $[L_\tau, U_\tau]$ (centré sur $\widehat{Y}_\tau$).

Question : intervalle et prévision pour $X_\tau = \exp(Y_\tau)$?

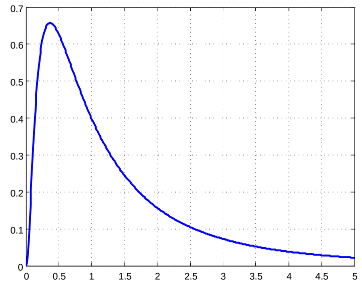
- Comme exp est croissante :
  $\Pr(\exp(Y_\tau) \in [\exp(L_\tau), \exp(U_\tau)]) \leqslant 95\%$.
  Donc « intervalle de confiance » au risque 5% :
  $[\exp(L_\tau), \exp(U_\tau)]$.
- Prévision ?
  Naïf : $\widehat{X}_\tau = \exp(\widehat{Y}_\tau)$
  Mieux : $\widehat{X}_\tau = \mathbb{E}(X_\tau) = \exp(\widehat{Y}_\tau + \widehat{\sigma}_\tau^2/2)$
  car X_τ suit une *loi log-normale*.

Illustration : $Y_t = \log(X_t)$

$\widehat{Y}_\tau = 0, \quad \widehat{\sigma}_\tau = 1.$



loi de Y_τ (normale)



loi de X_τ (log-normale)

Ici : $\widehat{Y}_\tau = 0$, I.C. risque 5% : $[-1.96, 1.96]$.
Prévision sur X_τ : I.C. risque 5% : $[0.14, 7.1]$
$\exp(\widehat{Y}_\tau) = 1$
$\widehat{X}_\tau = \exp(\widehat{Y}_\tau + \widehat{\sigma}_\tau^2/2) = 1.6$

Remarque : bien sûr, correction négligeable si $\widehat{\sigma}_\tau^2/2 << \widehat{Y}_\tau$

Tests de Dickey-Fuller

Objectif : test de stationnarité
On réécrit les processus AR + tendance :
$X_t = \alpha + \beta t + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t$
sous la forme :
$\Delta X_t = \alpha + \beta t + (\rho_1 - 1) X_{t-1} + \rho_2 \Delta X_{t-1} + \cdots + \rho_p \Delta X_{t-p+1} + \varepsilon_t$

Propriété : $(X_t - \alpha - \beta t)$ non stationnaire $\Leftrightarrow \rho_1 = 1$.

Les tests augmentés de Dickey-Fuller fournissent la p-valeur de $\widehat{\rho}_1/se$ ($\mathcal{H}_0 : \rho_1 = 1$) pour différentes valeurs de p dans les cas suivants :

- type 1 : $\alpha = \beta = 0$
- type 2 : $\alpha \neq 0$, $\beta = 0$
- type 3 : $\alpha \neq 0$, $\beta \neq 0$

En pratique : a priori on ne connaît ni le type, ni p...

$\rightarrow$ cf poly p. 58 (attention coquille $\boxed{\Delta} X_t = \ldots$ dans types 1,2,3)

Quelques « règles » complémentaires : différentiation (1)

- ordre de différentiation saisonnière : $D = 0$ ou $D = 1$.
- ordre de différentiation totale (saisonnière D & non-saisonnière d) : $d + D \leqslant 2$.
- si la décroissance de l'ACF est lente, penser à différentier plutôt qu'introduire un AR.
- si l'ACF est périodique (effet « pont suspendu »), alors différentiation saisonnière.
  $\rightarrow$ si X_t π-périodique, alors $\widehat{\rho}(h)$ aussi.

Quelques « règles » complémentaires : différentiation (2)

- si ACF pour décalage 1 est « franchement » négatif, la chronique est sans doute trop différentiée.

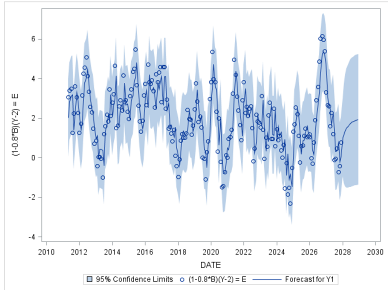
  $\rightarrow$ enlever un ordre de dérivation plutôt qu'introduire un MA.

  « *Justification* » :
  si X_t déjà stationnaire et $Y_t = (1-B)X_t$, alors :
  $\gamma_Y(1) = 2\gamma_X(1) - \gamma_X(0) - \gamma_X(2)$
  $\gamma_Y(0) = 2\gamma_X(0) - 2\gamma_X(1)$
  donc $\rho_Y(1) = -\frac{1}{2}\left(\frac{1-2\rho_X(1)+\rho_X(2)}{1-\rho_X(1)}\right)$
  et on est conduit à introduire un MA(1)...

Pourquoi éviter de trop différentier ?

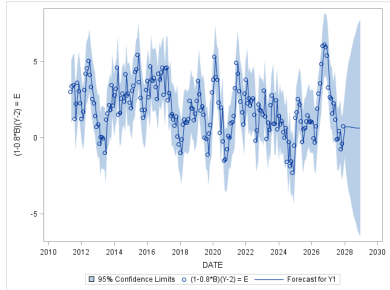
Cf intervalles de confiance de la prévision...

Exemple :



$(1-0.805B)Y_t = 0.3939 + \varepsilon_t$
$(\sigma = 1.002)$

ou $(1-0.805B)(Y_t - 2.02) = \varepsilon_t$



$(1-B)Y_t = -0.011 + \varepsilon_t$
$(\sigma = 1.052)$

Quelques « règles » complémentaires : la constante / moyenne

$$\Phi_p(B)\Phi_P(B^\pi)\left((1-B)^d(1-B^\pi)^D X_t - \underline{m}\right) = \Theta_q(B)\Theta_Q(B^\pi)\varepsilon_t$$

- chronique différentiée à l'ordre 1 :
  constante = pente de la tendance.
  On peut avoir une constante nulle
  (ex : marche aléatoire)
  ou pas
  (ex : marche aléatoire avec dérive, tendance « $at+b$ »)
  (pas très fréquent :...constante nulle par défaut dans `Arima`*)*

- chronique différentiée à l'ordre 2 (la pente « varie ») :
  constante = coef du terme quadratique.
  (tendance quadratique rare, donc *constante nulle* dans ce cas)

Quelques « règles » complémentaires : divers

- éviter de mélanger SAR et SMA.

- les termes en AR et MA peuvent se compenser.

  Ex : si ARIMA(2,d,1) identifié,
  on peut essayer ARIMA(1,d,0)
  (cas où les racines de AR et MA se « compensent »).

Sélection de modèle

Question : comment choisir entre différents modèles ARIMA(p,d,q) *valides* ?

Plusieurs possibilités :

- Minimisation de la variance des résidus σ^2

$\rightarrow$ mais plus le modèle est ≪ compliqué ≫ (p,d,q grands), plus σ est faible...

- Minimisation de *critères d'information* :

$\mathrm{AIC} = -2\log(\mathcal{L}) + 2k$ (Akaike)

ou $\mathrm{AICc} = \mathrm{AIC} + \frac{2k(k+1)}{T-k-1}$ (correction selon T)

ou $\mathrm{BIC} = \mathrm{SBC} = -2\log(\mathcal{L}) + k\log(T)$ (Schwarz Bayes)

$\rightarrow$ compromis entre vraisemblance $\mathcal{L}$ et nombre de paramètres estimés ($k = p + q + 1 + 0/1$ si ARIMA(p,d,q) complet)

Remarque 1 : ≪ le ≫ choix entre différents modèles se fait selon **un** critère.

Remarque 2 : on préfère les modèles simples (principe du *rasoir d'Ockham* ou de *parcimonie*).

Séance 3

Tous les modèles sont faux ... mais certains sont utiles !

"Remember that all models are wrong ; the practical question is how wrong do they have to be to not be useful."

George E. P. Box, Norman R. Draper
Empirical Model-Building and Response Surface,
Wiley, 1987.

Pour la prochaine séance : début de l'étude de la chonique `ozone`.