Journal of Quantitative Economics (2023) 21:1–40
https://doi.org/10.1007/s40953-022-00329-4

ORIGINAL ARTICLE

Multiscale Agricultural Commodities Forecasting Using Wavelet-SARIMA Process

Mamadou-Diéne Diop[1] · **Jules Sadefo Kamdem**[1,2]

Accepted: 27 September 2022 / Published online: 17 December 2022

Abstract

Forecasts of spot or future prices for agricultural commodities make it possible to anticipate the favorable or above all unfavorable development of future profits from the exploitation of agricultural farms or agri-food enterprises. Previous research has shown that cyclical behavior is a dominant feature of the time series of prices of certain agricultural commodities, which may be affected by a seasonal component. Wavelet analysis makes it possible to capture this cyclicity by decomposing a time series into its frequency and time domains. This paper proposes a time-frequency decomposition based approach to choose a seasonal auto-regressive aggregate (SARIMA) model for forecasting the monthly prices of certain agricultural futures prices. The originality of the proposed approach is due to the identification of the optimal combination of the wavelet transformation type, the wavelet function and the number of decomposition levels used in the multi-resolution approach (MRA), that significantly increase the accuracy of the forecast. Our SARIMA hybrid approach contributes to take into account the cyclicity and of the seasonality when predicting commodity prices. As a relevant result, our study allows an economic agent, according to his forecasting horizon, to choose according to the available data, a specific SARIMA process for forecasting.

Keywords Commodities · Forecast · Multi-resolution analysis · Wavelets · SARIMA

Introduction

Time series forecasting can be important for business and market decision support. It has been widely used, in particular for forecasting sales or for analyzing price variations from financial markets. Well-established forecasting methods are already

✉ Jules Sadefo Kamdem
jules.sadefo-kamdem@umontpellier.fr

[1] MRE UR 209 et faculté d'Économie, Université de Montpellier, Montpellier, France

[2] MRE UR MRE UR 209, (Université de Montpellier), Faculté d'Economie Avenue Raymond Dugrand, Site de Richter C.S. 79606, 34960 Montpellier Cedex 2, France

adopted by firms or market players, such as linear extrapolation and SARIMA. However, their performance remains limited since the time series studied are very volatile with some particular stylized facts. For instance, agricultural commodity prices sometimes present some specific stylized facts.

International institutions such as the World Bank and the International Monetary Fund, individual countries as well as companies involved in importing or exporting activities wish to forecasts prices of agricultural commodities or metals. Moreover, the World Bank expects a significant recovery in industrial commodities such as energy and metals in 2017, due to tighter supply and increased demand. With regard to certain energy or agricultural commodities, under certain assumptions, market forecasting publications are made in "Commodity Markets Outlook". These price forecasts allow the economic agent to identify the confrontation between supply and demand on a commodity market at a given future date. Predicting agricultural prices is difficult because other price series such as, the price of crude oil, the price of shares, or the prices of other financial assets, and the series of prices of agricultural commodities are influenced by several other uncertain abiotic factors (extreme weather variables, natural disasters, etc.) and biotic variables (pests, diseases, etc.) in addition to other invisible market forces and administrative measures (Wang et al. 2019). These explanatory factors taken together add more complexity to the analysis of the time series of prices agricultural products, which makes them difficult, an efficient forecast of agricultural prices. In recent literature (Xiong et al. 2018; Li et al. 2021) it's been shown that some vegetable price series are much more volatile and complex than the price series of other agricultural products due to their short duration and their seasonality. In addition, the perishable nature of vegetables further complicates obtaining effective vegetable price forecasts. The literature (Wang et al. 2019 and some references therin) highlights the complexities of price series, with as a corollary the difficulty of analyzing them in order to obtain a better forecast. Statistical models often used for forecasting agricultural prices include models like ARIMA (Box et al. 2015; Jadhav et al. 2018) and its constituent models (Hayat and Bhatti 2013). However, the use of the previous models does not take into account the heterogeneity of the agents involved in these agricultural markets. In this article, we propose a hybrid forecasting scheme that combines the classical SARIMA method and the wavelet transform (SARIMA-Wavelet). We believe that the proposed hybrid method is highly applicable for forecasting time series with specific stylized facts in the firm or the markets.

Also, the actors (arbitrageurs, hedgers or speculators) in the agricultural markets, do not have the same investment horizon. Thus, we opt in our paper to find the optimal SARIMA model, for each class of investors in the agricultural market, according to its investment horizon. To do this, in addition to the analysis of the complete series of available data, we carry out analyzes of the sub-series obtained from the available series, by means of a time-frequency decomposition of the available series, via recourse to wavelet theory. Intuitively, we know that the investment horizon is the reverse of the frequency. Thus, a frequency band is considered to be a band of investment horizons. We can therefore consider on each frequency band a time sub-series of the initial series. Thus, on this frequency band and therefore for this category of investors having investment horizons associated with the frequency band considered, it suffices to analyze the sub-series corresponding to this frequency band, to determine the optimal SARIMA process that will be used by the investors with a specific investment horizon. The procedure to build a SARIMA model on each sub-series consists of data preprocessing, model identification, parameter estimation, model diagnosis and finally application. For different sub-series, the optimal SARIMA model and it's parameters for the prices sub-series of both markets are shown.

Indeed, we define the best hybrid Wavelet-SARIMA approach for agricultural commodities price forecasting, which reflects quite clearly the fundamental concept of analysis in signal processing, where we decompose a complex and transitory signal into several sub-series. We highlight the pedagogical connection between the theory of wavelet transform and classical time series analysis SARIMA used in econometrics. A transitory signal is associated with a variable signal not periodic, which changes state suddenly. According to Yves Meyer[1], a wavelet is " the simplest transitory signal imaginable ". M. Misiti, Y. Misiti, G. Oppenheim, J-M. Poggi, in " *Wavelets and its applications* " (2003), define wavelets as " a signal processing tool for the analysis on several time scales, the local properties of complex signals can present areas of unsteadiness ". Therefore, analysis by wavelets help in the use of a well localized window fully scalable and along the signal to characterize the various components time-frequencies at any point. By the way, it is essential to use this method to identify the best statistical model able to describe each sub-series generated by the decomposition and give the best visibility of future values. For instance, the time-frequency Analysis of the Relationship Between EUA and CER Carbon Markets Sadefo Kamdem et al. (2016).

The wavelet transform is carried out in an amount of temporal subsets associated with frequency bands at the same sensitivity. In financial markets, each of the frequency bands represents a category of investors. Indeed, these agents adopt, according to the information they hold, a very heterogeneous behaviour. A combination of their actions produces very random changes in commodity prices. Thus, with the aim to support the regular decision making and monitor these markets through the introduction of prudent rules, the strategic conclusions of this document can help governments and economic authorities in the diagnosis and the detection of different speculative behaviour in the agricultural market.

In contrast to traditional Fourier analysis based on the frequency space, the wavelets analyse a signal in several horizons and frequencies by using the multiresolution analysis. These specificities are repeatedly solicited in many economic studies. For example: to identify the cyclical phenomenal changes in the market of stock indices, to study the co-movements and the effects of contagion between markets or within the same market. For that purpose, we base on the econometrics of stochastic processes in the time domain, but especially in that of frequencies using this theory. Indeed, it is interesting to understand the gaps between all investor behaviours as highlighted by their investment choices. However, in

[1] Yves Meyer is a Emeritus Professor at Superior Normal School of Cachan, Member of the Academic of Sciences since 1993. Specialist of harmonic analysis, he discovered the orthogonal wavelets.

economic times series with a lot of high frequencies, values are not in the same time interval, and thus it is not possible to apply the usual econometrics technical. Because, the applications of the appropriate methods are modelled for the database with the same interval, (cf. Engel and Russell, 1997– 1998). Hence we use the wavelet transform of random series where a signal projection is applied on analytical functions without any change in fundamentals properties. This allows us to highlight some characteristics of price and their variations: the hidden bumps and jumps detected during the evolution of a stock. They are usually caused by the impact of the few exogenous events not covered by the contract. In addition, there are seasonal features and extra-seasonal usually seen in a serial type namely the seasonal pattern that occurs permanently and regularly, the stochastic trend and / or seasonal phenomena that are cyclic. Finally, the volatility involved non-stationarity, the presence of any unit roots or phenomena long memories, non-linearity, etc. These points are important because they emphasize the essential information processed by wavelets.

The aim of this paper is to study the intriguing facts of combining the SARIMA model with wavelet transform to measure and anticipate the prices or returns time series. To achieve this, the descriptive analysis of Table 2 data, help to subtly explore the time series data of each product by studying its probability distribution.Some specificities detected thanks to the indicators of variability, asymmetry and fat tails of returns. This, to produce fair and reliable future values of monthly price indices. Three main points are given as follows:

- to test the best configuration of multi-resolution analysis by choosing one type of wavelet transform between (Cwt, Dwt and Modwt), the appropriate wavelet function and the number of decomposition on the robustness forecast of hybrid model Wavelet-SARIMA
- to calibrate the Wavelet-SARIMA model such that the combination selected in Mra (Mallat 1989) can realize good forecasts of price indices for cereals and oleaginous.
- to estimate the quality of forecast by the indicators like RMSE or MAE and compare it to classical SARIMA models and to White noise models on the basis of performance on series of price indices without any wavelet decomposition.

This is the best way to test significantly the effect of Mra configuration on economics times series forecasting using wavelet transform. Indeed, a large number of forecasts is simulated for having the smallest modelling error by solid tests.

The economic aspect of this paper is to contribute to the lighting of policy-makers and serve to aid decision-making tools of public policy or investment in the development control rules, sanctions and market security. These rules, once in place will serve as disincentives to speculative behaviour being adopted generally investors and therefore regulate and supervise the market for transactions of those raw materials (Fig. 1).

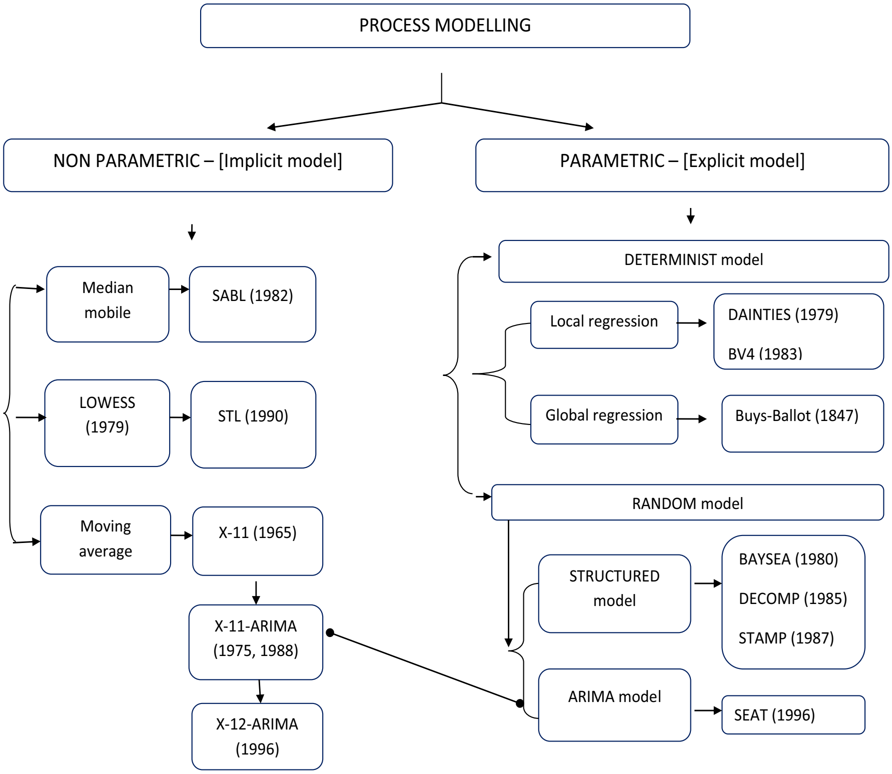


Fig. 1 Process modelling

Wavelet transform

The wavelet transform is a smart method capable of detecting all frequencies, and to consolidate those with the same sensitivity. Therefore, it separates all details or highs values from the trend in original times series.

The theoretical frame: Signal processing has greatly focused on the study of invariant operators in time or in space that modify properties of stationary signal. This has led to the reign of the Fourier transform, but leaving aside the essential of the information processing. To avoid losses, representations of time-frequency, have been developed to analyze any non-stationary process $(f(t))_{t\in T}$ indexed by t in any time space T using the transformation $W_f(u,s)$ (W as WAVELET) configured by two variables: the position u and the scale s. We consider $f(t)$ taking these values in $E\ \forall\ t \in T$:

$$T = \begin{cases} N, Z, R \ \ et \ \ E = R & \text{for univariate times series} \\ N, Z, R \ \ et \ \ E = R^d \ (d \geq 2) & \text{for multivariate times series} \end{cases}$$

A time-frequency representation is a transformation associating to $f(t)$ a real function of variables $W_f(u,s)$. It consist in a projection of signal on analysing functions $\psi_{u,s}$:

$$W_f(u,s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{+\infty} f(t)\overline{\psi_{u,s}(t)}dt \tag{1}$$

with :

- u the parameter of position, s the parameter of scale
- f the signal to analyze , $\psi_{u,s}$ the wavelet function chosen $\in L^2(R)$[2] and $\overline{\psi_{u,s}(t)}$ his conjugate. A wavelet ψ, defined in $L^2(R)$ presents at least the conditions below:

$$T = \begin{cases} \int_{-\infty}^{+\infty} |\psi(t)|dt = 0 \quad \text{vanishing moments} \\ \|\psi\| = \sqrt{\int_{-\infty}^{+\infty} |\psi(t)|^2 dt} = 1 \quad \text{the energy of analysing functions is constant} \end{cases}$$

The wavelets are regrouped by family. And, the most used in economics is the Daubechies family thanks to their excellent properties. From the single function ψ, we construct by translation (u) and by dilatation/contraction (s) a wavelets family representing the analysing functions.

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-u}{s}), \psi \; with \; s \in R^+, u \in R \tag{2}$$

The inverse wavelet transform helps to return an exactly reconstitution of the initial signal based on their coefficients of position and scale without any loss of information.

$$f(t) = \frac{1}{W_f(u,s)} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_f(u,s) \frac{1}{\sqrt{s}} \psi(\frac{t-u}{s}) \frac{dtds}{s^2} \tag{3}$$

The multi-resolution analysis: The fast algorithm of decomposition and reconstitution is applied to any process to view the representation of signals in different layers in order to have better visibility of local fluctuations at every stage of its resolution (Mallat 1989). The reconstitution is realized from the wavelets coefficients and scales via the inverse wavelet transform (equation 3). And, for more study, it will be possible to analyze and explain each frequency bands by spectral moments in a superior order, or doing forecast according to each sub-series generated by this algorithm. This type of analysis is important in financial markets, because it helps to decompose the evolution of their actions into several others signals. In the same vein, to identify potential flow accelerators per time horizon. However, it is more

[2] $L^2(R)$ is the set of square integrable functions: $\int_{-\infty}^{+\infty} |f(t)|^2 dt < +\infty$ and a Hilbert's space for the scalar product $\langle f, \psi_{u,s} \rangle$.

convenient to consider that it is the heterogeneous behaviour of agents that influences significantly the volatility of price indexes.

Doing the Mra need at first to choose a type of transform to do. With the discrete sample times series, we both apply a discrete wavelet transform (Dwt), a maximum overlap discrete wavelet transform (Modwt). The Dwt form determines a optimal number of wavelet coefficients and scales to decompose and reconstruct the signal. The Modwt form, uses all wavelet coefficients without any rebound possible. The contrary is the object of continuous wavelets transform (Cwt). It is very rebound and difficult to compute and to compile in practice. With these agricultural prices indices, we apply the additive decomposition based on the Modwt. Then, a list of wavelets is chosen, specially the Daubechies family. According to the signal to explore and it analysis, there are nowadays 25 wavelets functions with their properties. In economic and finance, the family of Daubechies is currently used. At the end, the number of decomposition is defined by $J \,\forall\, 2^J \leq N$ where N represent the number of data or sample.

$$f(t) = \sum_{j=1}^{J} D_{J,t} + A_{J,t} \tag{4}$$

The equation 4 is the result of the algorithm of decomposition-reconstitution at different scale j and calibrated on the Mra configuration. It separates the initial time series in a smooth series A_J into many other detail series $\left\{D_j \,\forall j = 1 \dots J\right\}$. The smooth image represents the general shape of signal in the half of its resolution. The details are all the high frequencies when we move from resolution j to $j+1$ ($[2^j - 2^{j+1}[$). During this period, the approximation is bigger. It grows bigger and bigger until all information is lost. But, by adding each detail D_j to A_J, the precision is increasingly brought in order to rebuild the original series. The table below defines $N\,/\,J\,/\,D_j\,/\,A_J\,/\,[2^j - 2^{j+1}]$.

Numerical data

Graphical representation

The Figs. 2 and 3 below represent respectively the changes of price indices of high agricultural products $I_t^{'}$ and their returns $\delta I_t^{'} = \frac{I_t^{'} - I_{t-1}^{'}}{I_{t-1}^{'}}$ between dates $t-1$ and t. It is about wheat, corn, sorghum, rice and oleaginous like soy, olive, palm and colza. Each chronicle has got 444 observations defined from January 1980 to December 2016[3]. To reduce the flexibility for modelling, a logarithmic transformation is applied on initial data $I_t^{'} = \log(I_t)$. Theoretically, the normalisation depend on the distribution of frequency values. The usual methodology developed by Box and Cox (1964) can help to give a better transformation by choosing the best γ parameter:

[3] The data time series are available on https://www.quandl.com.

$$I_t' = \begin{cases} \frac{I_t^{\gamma}-1}{\gamma} & \text{if} \quad \gamma \neq 0 \\ \log(I_t) & \text{if} \quad \gamma = 0 \end{cases}$$

Before the year 2000, changes seem to have sometimes the same behaviour. But, in 2007 and 2008, the price indices increase significantly. Indeed, strong spikes and troughs appeared in these years. According to an OCDE's report (2008) on the causes and consequences of agricultural commodities prices, these gaps are explained by a lot of combined factors. We notice the result of stationary production or inferior to the trend, a high increase of the demand and the investments on the derived agricultural product market. The exogenous factors are so complex. They are the periods of drought which affect the major grain areas, the weakness of the reserve of cereals and oleaginous, the development of use of agricultural commodities for biofuel production, the fast increase of crud oil prices. Finally, there is the continued devaluation of Us dollar, currency in which are generally expressed indicative prices of raw materials. All the changes intervened in a unstable context in the world economy, particularly the financial crisis of 2007/2008. This crisis effect on the speculative behaviour of producers or financial agents. Moreover, the rate of increase represented in parallel give more visibility on the flexibility of price indices. Because, a surge is often offset by a smaller increase in the following month, hence the appearance of those many " peaks and troughs ".

In the Figs. 2 and 3 it is complex to identify clearly at first, a seasonal and a trend in any data series. As an alternative, we aggregate annually (Fig. 11 in Appendix) and monthly (Fig. 12 in Appendix) in order to highlight these general characteristics or other cyclical effects. On these graphics, there are " peaks and troughs " particularly with corn data in different periods. For the soy and the sorghum, " peaks " are in November and December and the " troughs " from June to August. The corn increase more and more until it " peaks " (June), then decreases to its poor levels in September to leave relatively well until the end of the year. The rice has the same changes but at different extremes: peaks " on April and " troughs " on the last trimester. The series of soy and olive changes almost together but with a noticeable difference in the second half. So, their maximum are achieved respectively in November and August. As to palm and colza series, we notice the same trajectory: a strong start and a fall from May and June. But we detect no seasonality. However, at the annual average prices, a periodic behaviour and a trend seems to be emerging over time (Table 1).

Descriptive statistics

The Table 2 present a descriptive statistics summary on data. It highlights some stylized facts of these data. According to the coefficients of variation ($\frac{\sigma}{mean}$), we can see a high variability in the price times series.

Table 1 MRA in different level of scale

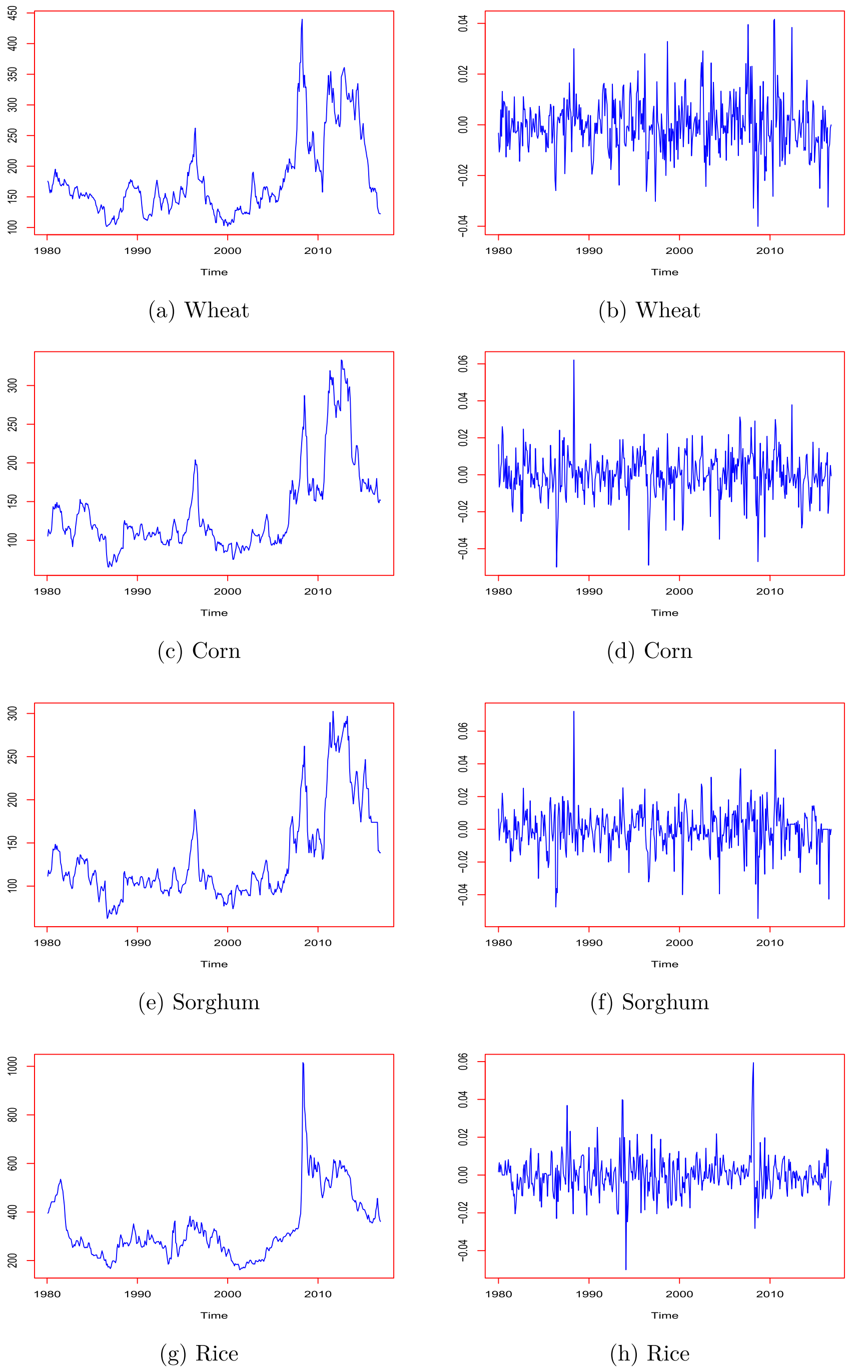
Observation N	Level J	Sub-series $D_1, \ldots, D_J, A_J$	Resolution
[100–127]	6	D_1	[2–4[
		D_2	[4–8[
		$\vdots$	$\vdots$
		D_6	[64–128[
		A_6	[64–128[
[128–255]	7	D_1	[2–4[
		D_2	[4–8[
		$\vdots$	$\vdots$
		D_7	[128–256[
		A_7	[128–256[
[256–511]	8	D_1	[2–4[
		D_2	[4–8[
		$\vdots$	$\vdots$
		D_8	[256–512[
		A_8	[256–512[
[512–1023]	9	D_1	[2–4[
		D_2	[4–8[
		$\vdots$	$\vdots$
		D_9	[512–1024[
		A_9	[512–1024[

High volatility reflects a lot of information hidden: the high irregularity of price indices, their non stationary, an unstable market, the impacts of exogenous factors or random phenomena. They are impossible to explore by a simple econometric model. The skewness $\beta_1^{1/2} = \frac{\mu_3}{\mu_2^{3/2}}$ and kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2}$ coefficients are indicators of asymmetry and fat tails of price returns. So, if they follow a normal distribution then $\beta_1^{1/2} = 0$ and $\beta_2 = 3$[4], but that is not the case. Indeed, the positive skewness demonstrates an asymmetric distribution to the right due to the impact of extremes values. The kurtosis are bigger than 3 ($\beta_2 > 3$), so more concentrated in opposite to the normal distribution.

Processing and data analysis

We try to find the best SARIMA model for explaining the increase rates of agricultural price indices and make forecast. So, the data are decomposed in two periods: from January 1980 to December 2016 (37 years / 444 observations) and the 2017 period. The first part is a framework of calibration of econometric models.

4 symmetric distribution and a flattened like Gauss's.



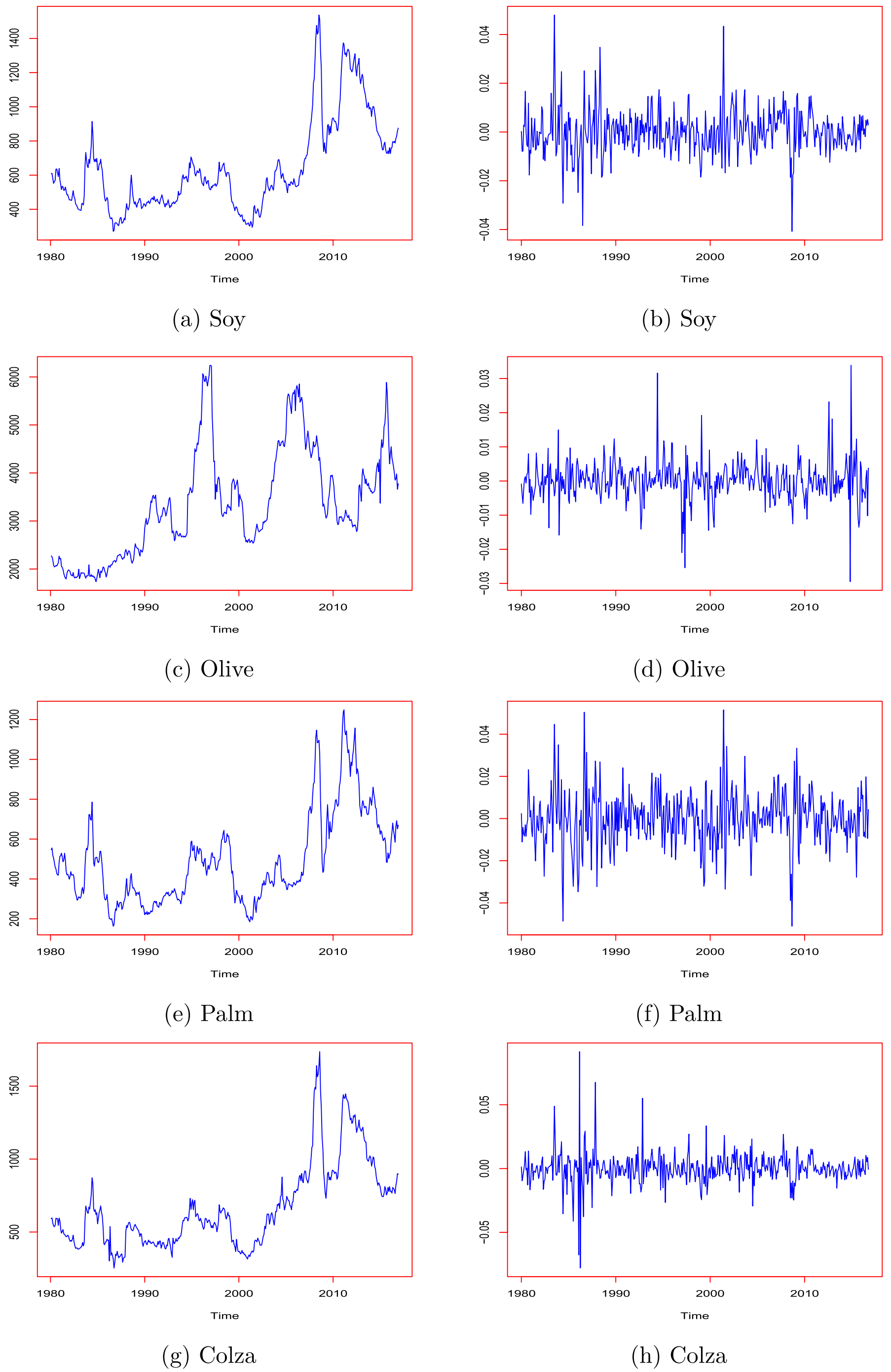
(a) Wheat (b) Wheat

(c) Corn (d) Corn

(e) Sorghum (f) Sorghum

(g) Rice (h) Rice

Fig. 2 Indices and returns cereals prices



(a) Soy (b) Soy

(c) Olive (d) Olive

(e) Palm (f) Palm

(g) Colza (h) Colza

Fig. 3 Indices and returns oils prices

Table 2 Basics descriptive statistics for times series data

	Mean	Median	Std. dev	Coef. var	Kurtosis	Skewness	Normality (%)
Wheat	178.95	159.25	65.57	0.37	4.38	1.42	< 5
Corn	137.40	115.35	56.96	0.41	5.20	1.64	< 5
Sorghum	132.74	111.64	53.22	0.40	4.20	1.42	< 5
Rice	334.81	292.00	134.68	0.40	5.66	1.42	< 5
Soy	639.30	563.00	268.17	0.42	3.97	1.25	< 5
Olive	3349.92	3164.79	1124.33	0.34	2.61	0.60	< 5
Palm	493.09	439.74	224.66	0.46	3.83	1.11	< 5
Colza	666.33	590.66	290.51	0.44	4.15	1.24	< 5

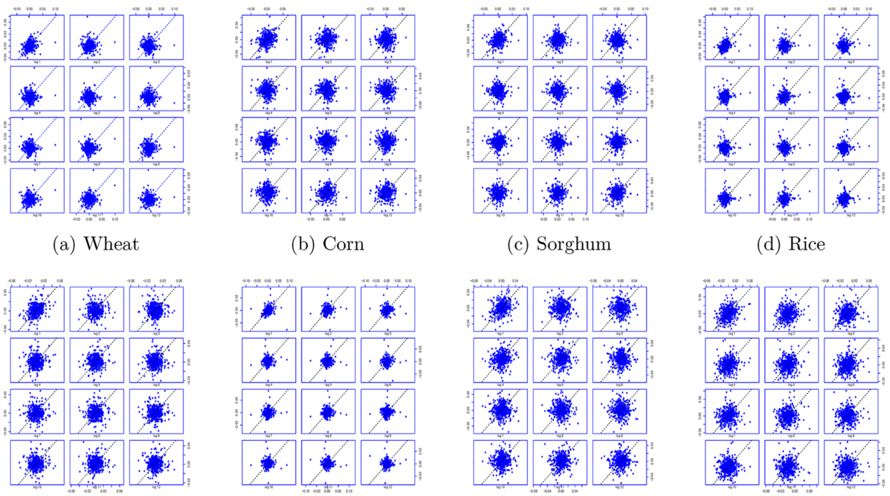
Table 3 Stationary

	P_{value} (%)	Dickey-Fuller	P_{value} (%)	Philips-Perron	P_{value} (%)	Kpss
Wheat	< 5	Reject	< 5	Reject	< 5	Accept
Corn	< 5	Reject	< 5	Reject	< 5	Accept
Sorghum	< 5	Reject	< 5	Reject	< 5	Accept
Rice	< 5	Reject	< 5	Reject	< 5	Accept
Soy	< 5	reject	< 5	Reject	< 5	Accept
Olive	< 5	Reject	< 5	Reject	< 5	Accept
Palm	< 5	Reject	< 5	Reject	< 5	Accept
Colza	< 5	Reject	< 5	Reject	< 5	Accept

We built the SARIMA model by studying the fundamental characteristics (seasonality, stationary, cyclical phenomenal, unit root, trend). Indeed, we propose a model, estimate the parameters, test its suitability and analyse the residuals. Several models can intervene or agree. But thanks to the selection criterion AIC or Bic and especially with sparingly, simple models are preferred. Then, the model chosen is used to re-estimate the 2017 values. The indicators below are used to measure the accuracy of the forecast:

- mean absolute error $MAE = \frac{1}{n}\sum_1^n \left|x_t - \hat{x}_t\right|$
- root mean error $RMSE = \sqrt{\frac{1}{n}\sum_1^n \left(x_t - \hat{x}_t\right)^2}$

Diagnostics: Based on the returns data series $\delta I_t'$, we find the number of lags that are able to denoise the residuals. Indeed, we start with a general modelling with a linear trend (Trend) and respectively with a drift (Drift) and none (None) model. On each case, we apply the unit root test for stationarity according to Dickey-Fuller (1979). But, Given the annual seasonality observed in the qualitative analysis (Fig. 11 in Appendix) and empirical (Table 4), we choose a lag 11. After



(a) Wheat (b) Corn (c) Sorghum (d) Rice

(e) Soy (f) Olive (g) Palm (h) Colza

Fig. 4 Lagplot of returns times series

estimation, the maximum lag is 0 instead of 11. In addition, we use other tests of Philips-Perron (1987) and Kpss (1992) in Table 3.

The p_{values} of three models are all inferior to 5% (Table 3). It means there is stationarity in returns data series. It is slightly visible on Fig. 2 and 3 in "Numerical data") even if there are some spikes and troughs. Indeed, the calculations of returns removes any trend in these series to focus mainly on seasonal effects and the rest we hope to be white noise or modelling a stationary process. So, it is not necessary to make them stationary. In addition, the graphics of lags (Fig. 4) present no linear relationship. It opposes the returns series by themselves shifted some lag ($\forall\ h = 1 \ldots 12$). If the point clouds along the right equation $y = x$, then there is a strong autocorrelation else a dispersion around the means. No graphics supports the presence of trend.

The empirical test has been implemented to determine the number of differentiation required for having stationarity. There are for example the test of Ocsb: (Osborn et al. (1988) and that of Hyndman and Khandakar (2008). The result " 1 " means that there exists a seasonal unit root contrary to " 0 " (Table 4). Moreover, the methodology highlights the *auto.arima* function in R and by using their package forecast helps us to verify the seasonality. It is inspired by precursors like Hannan and Rissanen (1982), Liu (1989), Gomez and Maravall (1998), Melard and Pasteels (2000). It starts from the general model $(Arima(p, d, q)(P, D, Q)s$ to recover the degree of differentiation (the parameter I of Arima) providing seasonality.

The seasonal adjustment of returns times series $\delta I'_t$ is done thanks to a linear filter F at lag 12 like: $F = 1 - B^{12}$. We obtain a new time series $S_t = \delta I'_t(1 - B^{12})$.

Framework of calibration: Arima is a generalization of an Auto-Regressive Integrated Moving Average model that represents an important example of the Box and

Table 4 Seasonality

	Monthly seasonal	Test	Annual seasonal	Test
Wheat	0	Reject	1	Accept
Corn	0	Reject	1	Accept
Sorghum	0	Reject	1	Accept
Rice	0	Reject	1	Accept
Soy	0	Reject	1	Accept
Olive	0	Reject	1	Accept
Palm	0	Reject	1	Accept
Colza	0	Reject	1	Accept

Jenkins (1976) approach. In particular, the SARIMA model contains a seasonal component and it is the famous linear model for time-series analysis and forecasting. With its success, it occupies an essential place in academic research and in fields such as economics, finance and agro-industry. A time series X_t $\forall\, t = 1, 2, \dots, k$ is generated by a $Sarima(p, d, q)(P, D, Q)_s$ process if:

$$\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D X_t = \theta_q(B)\Theta_Q(B^s)\epsilon_t \tag{5}$$

where N is the number of observations; p, d, q, P, D and Q are positives integers; B is the lag operator; s is the seasonal period length.

- d reflects the initial differentiation obtained with the calculation of increases in rates in the times series between the dates $t-1$ and t by $\delta I'_t = \frac{I'_t - I'_{t-1}}{I'_{t-1}}$. It means the number of regular differences is ($d <= 2$) (Shumway and Stoffer 2006): so $d = 1$.
- s represents the annual seasonal adjustment observed and determined above respectively at Fig. 12 in Appendix: so we have $s = 12$ and $D = 1$. D is the number of seasonal differences. If there is seasonality effect, $D = 1$ in the most cases. Else, if there is no seasonality effect, $D = 0$.
- ϵ_t is the estimated residual at time t that is identically and independently distributed as a normal random variable with an average value equal to zero $\mu_\epsilon = 0$ and a variance σ_ϵ.
- The orders p and P are the parameters of the regulars seasonal and autoregressive operator (AR and SAR). They are determined by using the partial auto-correlogramme function FAP.
- The orders q and Q are those of the regulars seasonal and moving average functions (MA and SMA), they are determined by using the simple auto-correlogram function FAC.

$$Sarima \begin{cases} \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p & \text{is the Ar(p)} \\ \Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps} & \text{is the Sar(p)} \\ \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^q & \text{is the Ma(q)} \\ \Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_p B^{ps} & \text{is the Sma(Q)} \end{cases}$$

Autoarima model

At first, we use the AUTOARIMA model developed and implemented by Hyndman et al. (2008) before using the equation 5. It built a better model with the best AIC or BIC criterion. The terms of errors ϵ_i are estimated by the residuals $r_i = y - \hat{y}$. We suppose they are a white noise signal following a normal distribution $N(0, \sigma^2)$. So: $E(\epsilon_i) = 0 \ \forall\ i$ and $V(\epsilon_i) = \sigma^2 \ \forall\ i$). The variance of the residuals from the model is not homogeneous, but depend on the position of observation i. Thus, we check for stationary behaviour by plotting the residuals r_i on times t. The independence between ϵ_i is managed by $\frac{cov(r_i,r_j)}{\sqrt{s^2(r_i)}\sqrt{s^2(r_j)}} = 0$ for any point (i, j). For verification, the Durbin-Watson or Ljung-Box test is use.

The graphics of validation model[5] help to test the efficiency of each model via its residuals. The residuals graphics r_i in term of times is generally the first diagnostic. Indeed, thanks to its logical order in having data, these representations support the missing or not of serial positive correlations between ϵ_i. The auto-correlations functions and its probabilities of Ljung-Box test support that the residuals are all auto-correlated. Indeed, on the graphics, some points are out of the reject area of missing auto-correlation of ϵ_i. The double differentiation at the orders 1 and 12 on original data series probably did not completely rule out the dependence between observations. An empirical test confirm the presence of auto-correlations between the r_i according to their weak P_{value} ($P_{value} < 0,05$). But, a focus on the simple auto-correlation, show the lags superiors to 12 are slightly higher and thus reveal a slight long-term dependency.

Seasonal Arima model

The White noise model is the simplest modelling because, it takes no auto-regressive and no moving average component.

The first autocorrelation coefficients are higher and always support a short-term dependency. Following the modelling, we add in the moving average and auto-regressive components respectively on the SMA et SAR parts from the white noise model. This methodology built the best SARIMA model. At first, we start by SMA a model like $Sarima(0, 1, q)(0, 1, Q)_{12}$. The orders q and Q are determined by using the simple autocorrelation function FAC on stationary time series. For the order q, we look for the last lag which leaves the band at the same time inferior to 12. For the order Q, the selection is done on all the lag multiples to 12 which also leave the same band. So after analysis, we are:

q orders: $q_{Wheat} = 10,\ q_{Corn} = 2,\ q_{Sorghum} = 10,\ q_{Rice} = 8,\ q_{Soy} = 10,\ q_{Olive} = 3,$ $q_{Palm} = 11$ and $q_{Colza} = 10$.

[5] graphics of standardised residual, graphic of simple and partial autocorrelation function and graphic of Ljung-Box test.

Q orders: $Q_{Wheat} = 1,\ Q_{Corn} = 2,\ Q_{Sorghum} = 1,\ Q_{Rice} = 2,\ Q_{Soy} = 1,\ Q_{Olive} = 1,$ $Q_{Palm} = 2$ and $Q_{Colza} = 1.$

Now, the modelling is about SAR model like $Sarima(p, 1, 0)(P, 1, 0)_{12}$. From the White noise model, we complete by adding the news components on the seasonal and auto-regressive part. Based on the partial autocorrelation function of stationary time series, we applied the rule above to choose the orders p and P.

p orders: $p_{Wheat} = 10,\ p_{Corn} = 1,\ p_{Sorghum} = 11,\ p_{Rice} = 9,\ p_{Soy} = 11,\ p_{Olive} = 10,$ $p_{Palm} = 11$ and $p_{Colza} = 11.$

P orders: $P_{Wheat} = 2,\ P_{Corn} = 2,\ P_{Sorghum} = 2,\ P_{Rice} = 2,\ P_{Soy} = 2,\ P_{Olive} = 2,$ $P_{Palm} = 2$ and $P_{Colza} = 2.$

In this part, we combine the MA and SARI components calculated above in order to find a better SARIMA model like $Sarima(p, 1, q)(P, 1, Q)_{12}$ for all agricultural data. In the precede modelling, there could be some complexity because of their very high orders p, P, q and Q, even if their AIC criterion have improved significantly compared to previous models. By parsimony, we have shown and set to 0 insignificant coefficient to simplify. We calculated the p_{value} of estimated parameters and analyze the models and remove those whose probabilities are superior than 0.05. This process is thus repeated until we obtain the best models where all coefficients are significantly non-zero.

On the graphics of Ljung-Box (Figs. 13 and 14 in Appendix) some points are out of the band to reject with autocorrelation missing. In contrast, the graphics of simple autocorrelation function show almost no significant lag. The residuals r_i are stationary. Besides, on the representations of standardized residuals, the graphics reflect an average behaviour of the evolution of the mean and variance over time.

These models are more precise for econometric analysis and they take into account the singular and regular parts of S_t. Furthermore, they represent the best AIC, MAE and RMSE for forecast and oppose to the hybrid WAVELET-SARIMA model.

Hybrid forecasting based on wavelet transform and Sarima

The processing WAVELET-SARIMA model requires at first, an optimal MRA setting. Then, we define a type of transform, the wavelet function and the number of decomposition to use on those data series.

Table 5 Times series analysis with Autoarima model

	Model	Aic	Kpss Test	Ljung Test	Rmse	Mae
Wheat	Sarima(0,0,1)(0,1,0)[12]	4165.49	0.10	0.00	30.02	20.96
Corn	Sarima(1,0,0)(0,1,0)[12]	3483.11	0.10	0.00	13.66	8.86
Sorghum	Sarima(1,0,0)(0,1,0)[12]	3394.98	0.10	0.00	12.34	8.37
Rice	Sarima(5,0,3)(0,1,0)[12]	4192.58	0.10	0.25	30.41	17.53
Soy	Sarima(0,0,1)(0,1,0)[12]	5266.27	0.10	0.00	108.12	76.13
Olive	Sarima(0,0,2)(0,1,0)[12]	6268.66	0.10	0.00	345.22	244.14
Palm	Sarima(0,0,1)(0,1,0)[12]	5178.87	0.10	0.00	97.26	70.88
Colza	Sarima(2,0,5)(0,1,0)[12]	4830.2	0.10	0.85	64.25	45.96

Table 6 Times series analysis with White noise model

	Model	Aic	Kpss Test	Ljung Test	Rmse	Mae
Wheat	Sarima(0,1,0)(0,1,0)[12]	3748.58	0.10	0.00	18.87	11.90
Corn	Sarima(0,1,0)(0,1,0)[12]	3484.02	0.10	0.00	13.87	8.87
Sorghum	Sarima(0,1,0)(0,1,0)[12]	3396.60	0.10	0.00	12.53	8.33
Rice	Sarima(0,1,0)(0,1,0)[12]	4377.23	0.10	0.00	39.19	20.62
Soy	Sarima(0,1,0)(0,1,0)[12]	4725.05	0.10	0.00	58.73	40.48
Olive	Sarima(0,1,0)(0,1,0)[12]	5953.86	0.10	0.00	245.14	161.69
Palm	Sarima(0,1,0)(0,1,0)[12]	4704.01	0.10	0.00	57.31	40.51
Colza	Sarima(0,1,0)(0,1,0)[12]	4881.21	0.10	0.00	70.43	47.30

$$Mra = \begin{cases} Dwt & / \; Wavelet \; / \; J \\ Modwt & / \; Wavelet \; / \; J \\ Cwt & / \; Wavelet \; / \; J \end{cases}$$

The data analysis are applied on each sub-series in order to identify the best model able to describe the speculative behaviour adopted by people in order to monetize its kitty. At the end, the global forecast is obtained by adding the forecast of new sub-series. The application of this hybrid methodology is used only in two data series: wheat and soy data series. We hope that it can improve the accuracy in the forecast better than the Sarima model (Tables 5, 6, 7, 8, 9).

Configuration of multiresolution analysis

The wheat data sub-series and it variation represented in Figs. 5 and 6 are the result of the multi-resolution analysis. These data sub-series are separated into a first component smooth (the general appearance) and a set of details base on their resolution $[2^j;2^{j+1}[$. Each frequency band is specific because it describe the change of wheat's

Table 7 Times series analysis with Sma model

	Model	Aic	Kpss Test	Ljung Test	Rmse	Mae
Wheat	Sarima(0,1,10)(0,1,1)[12]	3416.86	0.10	0.94	11.88	7.90
Corn	Sarima(0,1,2)(0,1,2)[12]	3184.07	0.10	0.91	9.20	5.89
Sorghum	Sarima(0,1,10)(0,1,1)[12]	3135.20	0.10	0.94	8.57	5.58
Rice	Sarima(0,1,8)(0,1,2)[12]	4022.45	0.10	0.95	24.12	13.52
Soy	Sarima(0,1,10)(0,1,1)[12]	4384.58	0.10	0.99	36.63	25.16
Olive	Sarima(0,1,3)(0,1,1)[12]	5708.92	0.10	0.98	178.51	114.76
Palm	Sarima(0,1,11)(0,1,2)[12]	4348.91	0.10	0.96	34.94	23.19
Colza	Sarima(0,1,10)(0,1,1)[12]	4591.01	0.10	0.99	46.58	31.52

Table 8 Times series analysis with Sma model

	Model	Aic	Kpss Test	Ljung Test	Rmse	Mae
Wheat	Sarima(10,1,0)(2,1,0)[12]	3512.90	0.10	0.95	13.82	9.33
Corn	Sarima(1,1,0)(2,1,0)[12]	3281.31	0.10	0.74	10.78	6.89
Sorghum	Sarima(8,1,0)(2,1,0)[12]	3192.92	0.10	0.95	13.82	6.34
Rice	Sarima(4,1,0)(2,1,0)[12]	4115.71	0.10	0.99	28.32	16.87
Soy	Sarima(4,1,0)(2,1,0)[12]	4478.13	0.10	0.92	43.12	30.13
Olive	Sarima(4,1,0)(2,1,0)[12]	5752.73	0.10	0.88	188.91	126.75
Palm	Sarima(8,1,0)(2,1,0)[12]	4498.02	0.10	0.97	43.98	31.28
Colza	Sarima(8,1,0)(2,1,0)[12]	4749.66	0.10	0.95	58.97	41.66

Table 9 Times series analysis with Sma model

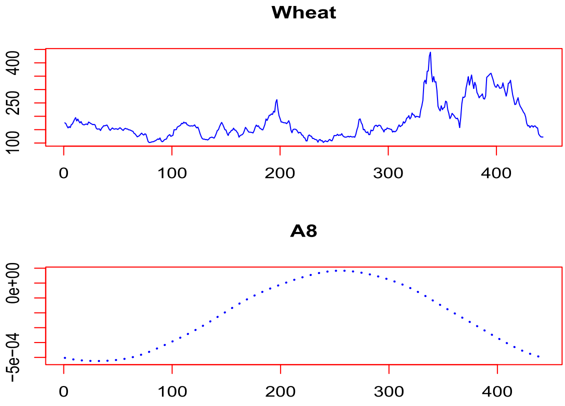
	Model	Aic	Kpss Test	Ljung Test	Rmse	Mae
Wheat	Sarima(0,1,2)(2,1,1)[12]	3420.10	0.10	0.97	12.05	7.76
Corn	Sarima(0,1,2)(0,1,2)[12]	3185.07	0.10	0.91	9.20	5.89
Sorghum	Sarima(1,1,4)(0,1,2)[12]	3149.94	0.10	0.99	8.82	5.73
Rice	Sarima(2,1,4)(0,1,2)[12]	4026.89	0.10	0.96	24.39	13.47
Soy	Sarima(2,1,1)(0,1,2)[12]	4381.21	0.10	0.99	36.95	25.27
Olive	Sarima(1,1,2)(2,1,1)[12]	5712.48	0.10	0.99	178.42	114.70
Palm	Sarima(2,1,1)(0,1,1)[12]	4369.32	0.10	0.87	36.58	25.60
Colza	Sarima(2,1,1)(0,1,2)[12]	4590.84	0.10	1.00	47.17	31.66

price indices in any horizon. So, there are the high, middle and low frequencies. In the agricultural market, each class of agents or producers acts according to their investment horizons. The same multi-resolution analysis is done with soy data series and his variation (cf. Figs. 15 and 16 in Appendix). In both case, working with the variation data series by using this kind of resolution is more interesting in multi-scale analysis.

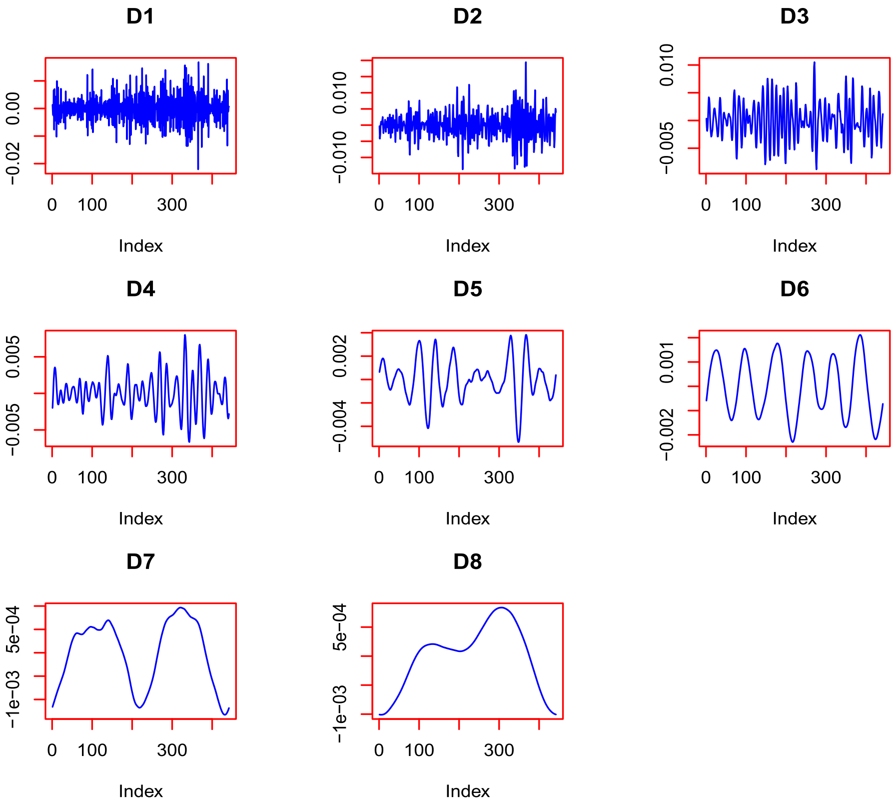
Table 10 Basic descriptives statistics for times sub-series

	D1	D2	D3	D4	D5	D6	D7	D8	A8
Wheat									
Kurtosis	8.71	7.95	4.18	8.10	5.31	2.60	2.00	1.81	1.50
Skewness	0.10	0.13	0.09	0.84	0.13	0.29	0.27	0.64	0.02
Shapiro Test	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Kpss Test	0.10	0.10	0.10	0.10	0.10	0.10	0.01	0.01	0.01
Nb.diffs	0	0	0	0	0	0	2	2	2
LogWheat									
Kurtosis	3.56	4.55	2.83	3.57	3.47	1.75	1.76	2.11	1.49
Skewness	0.03	0.04	0.07	0.17	-0.23	-0.10	-0.44	-0.39	-0.02
Shapiro Test	0.12	0.00	0.66	0.00	0.00	0.00	0.00	0.00	0.00
Kpss Test	0.10	0.10	0.10	0.10	0.10	0.10	0.02	0.01	0.01
Nb.diffs	0	0	0	0	0	0	2	2	2
Soy									
Kurtosis	10.03	5.79	10.74	13.41	4.15	2.27	2.22	1.76	1.50
Skewness	0.17	0.40	-0.11	1.30	0.16	0.25	0.40	0.59	0.01
Shapiro Test	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Kpss Test	0.10	0.10	0.10	0.10	0.10	0.10	0.01	0.01	0.01
Nb.diffs	0	0	0	0	0	0	0	2	2
LogSoy									
Kurtosis	4.33	5.46	4.13	6.15	2.69	2.65	2.03	2.06	1.51
Skewness	0.04	0.11	-0.14	-0.30	0.04	-0.23	-0.04	0.15	0.01
Shapiro Test	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Kpss Test	0.10	0.10	0.10	0.10	0.10	0.10	0.06	0.01	0.01
Nb.diffs	0	0	0					2	2

For this way, we analyse to find the best modelling for any agricultural commodities. The Table 10 uses the same indicators like in Table 2. It specifies information by time horizon. The differences between the measurements indicate that the first frequency bands have more memory than the others. In addition, they decrease when the space-time become larger. The kurtosis being slightly superior than 3 supports the presence of a low leptokurticity compared to normal Gauss distribution. The skewness are positives or negatives according to sub-series. The KPSS's test determine the missing of unit root in bands of wheat and soy data series. The probabilities of this test are inferior than 5% in $D(7)$, $D(8)$ and $A(8)$ band. But, for the other sub-series: $D(1)$, $D(2)$, $D(3)$, $D(4)$, $D(5)$ and $D(6)$, the $P_{value} = 10\%$. It means that there is no unit root. So they are stationary in opposite to those with a low frequency unless we use a double differentiation in modelling.

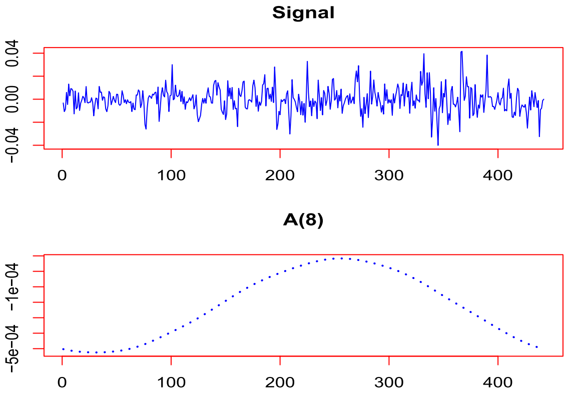


(a) Approximation

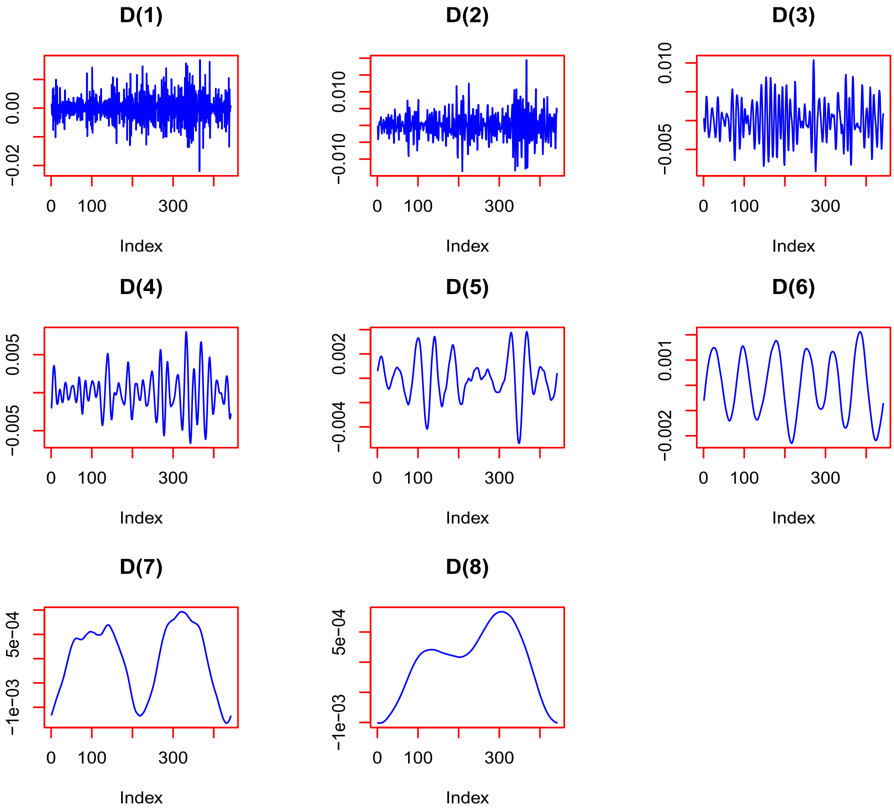


(b) Details

Fig. 5 MRA on Wheat times series



(a) Approximation



(b) Details

Fig. 6 MRA on LogWheat times series

Framework of calibration and validation

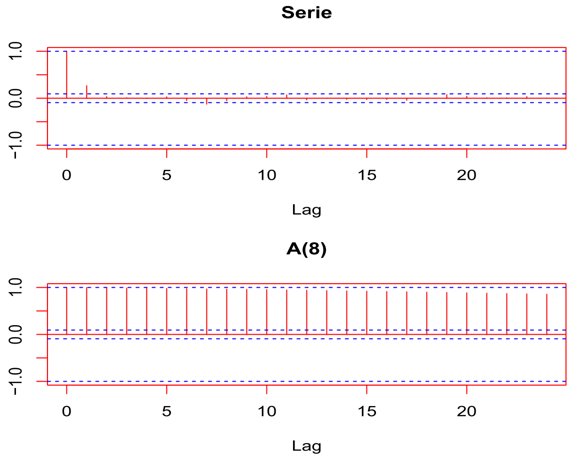
Each sub-series is modelled by using directly information obtained with the simple and partial autocorrelation function (Fac and Fap). But before modelling, an empirical test for unit root on the bands denote the low probabilities test ($\leq$ 0;05) for $D7$, $D8$ and $A8$, so no stationary and we must make them stationary. They must be neutralized by a filter $F = (1 - B)^2$. In contrast, the bands $D1$ to $D6$ are stationary. Regular process requiring the knowledge of an infinite number of parameters to specify them fully. We choose the SARIMA process which represent a huge class of stationary processes and needs some parameters. The Fac's and Fap's graphics are used to define the orders q, Q, p, P.

Simple and Partial autocorrelation function: The Fac's Figs. 7a and b for wheat and 9a and b for soy are many lags out of the critical area and they are different. If one considers the high frequencies of the signals $D1, \ldots D6$, we identify the orders q and Q in modelling. For the bands $D7$, $D8$ and $A8$, a differentiation is necessary. Indeed, some bands of each data prices highlight all characteristics omitted and allow the detection of high values, seasonal, cyclical and trend effect. The Fap's Figs. 8a and b for wheat and 10a and b for soy, are important lags depending on the time horizon. The orders p and P of auto-regressive and their seasonal version parts are important in high frequencies and decrease at low frequencies. When, they are no information in $D7$, $D8$ and $A8$, the move is similar to an auto-regressive process at order 1 (Ar(1)).

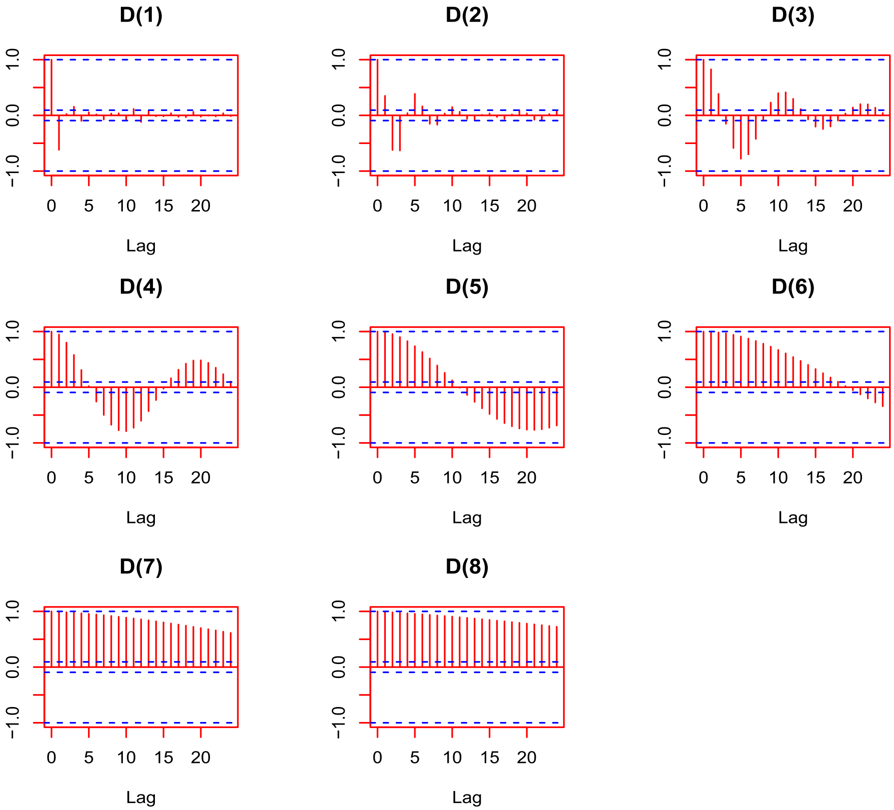
Diagnostic and validation: The normality test have probabilities inferior to 0.05. So the normal of residuals hypothesis is to reject. The stationarity test and the missing of auto-correlation between the terms of errors. We use the test of KPSS and Ljung-Box. The P_{value} are higher than 0.05 on all frequencies. The residuals are therefore stationary and no auto-correlated. The Table 11 represented all informations detected in the residuals diagnostic and the quality of forecast.

Discussions and conclusion

According to the difference modelling, the complete Seasonal ARIMA model $Sarima(p, d, q)(P, D, Q)_{12}$ is better in the forecast of agricultural data series. Thus, it'll be used for forecasting the futures values. However, the tool of multi-resolution analysis of wavelet theory is more specific because it carries a segmented analysis of chronic resting on exploration temp and scale. t identifies that these time series are further broken down into components although the trend and seasonality. Moreover, the same results prove that the Wavelet-Sarima model give the better accuracy with the root mean square error *RMSE* both its calibration as its validation. The decomposition of these series in itself is already a first resolution of the complexity hidden information. Thanks to the strength of wavelets, the modelling improved greatly the usual models. This precision and technical ability provided by this model combinations $Wavelet - Sarima$ / $Wavelet - Arima$ are demonstrated by Conejo et al. (2005)

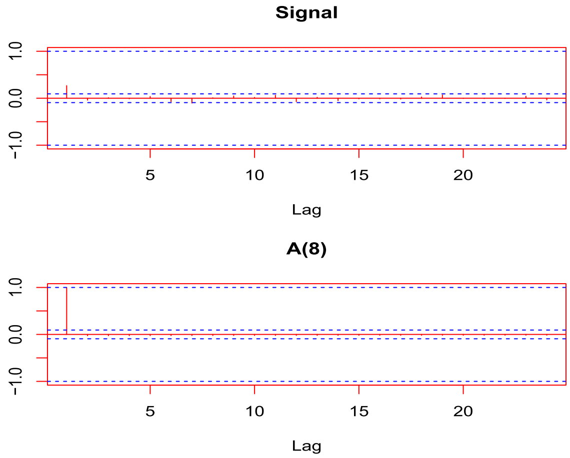


(a) Approximation

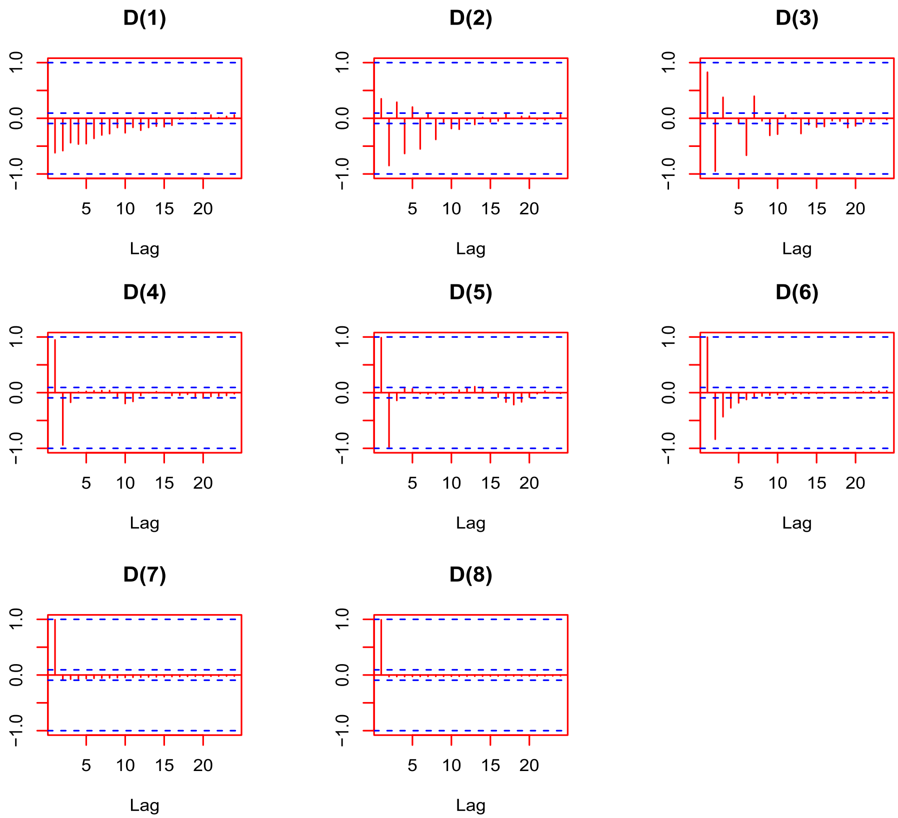


(b) Details

Fig. 7 Wavelet Simple autocorrelation function - LogWheat

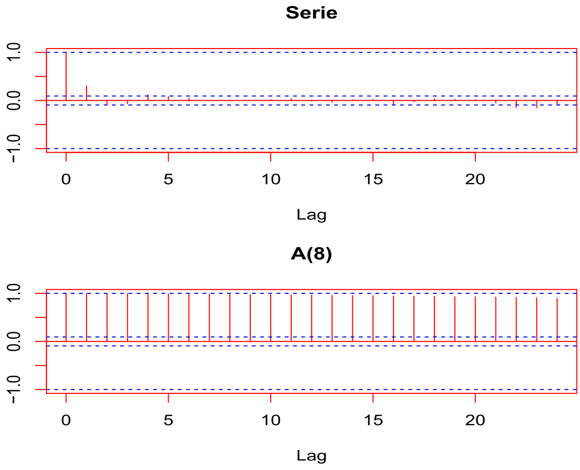


(a) Approximation

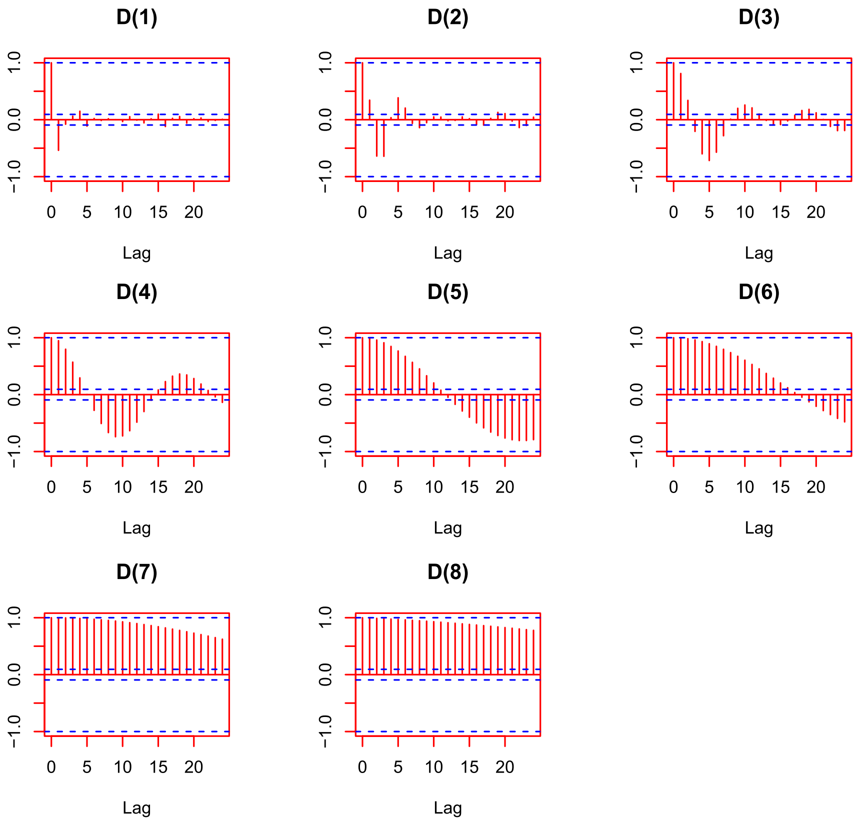


(b) Details

Fig. 8 Wavelet Partial autocorrelation function - LogWheat

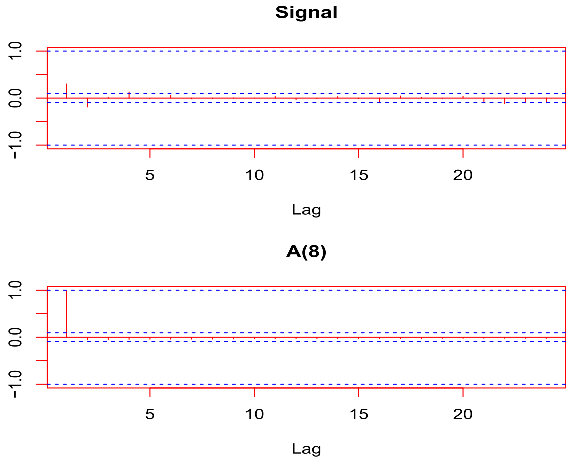


(a) Approximation

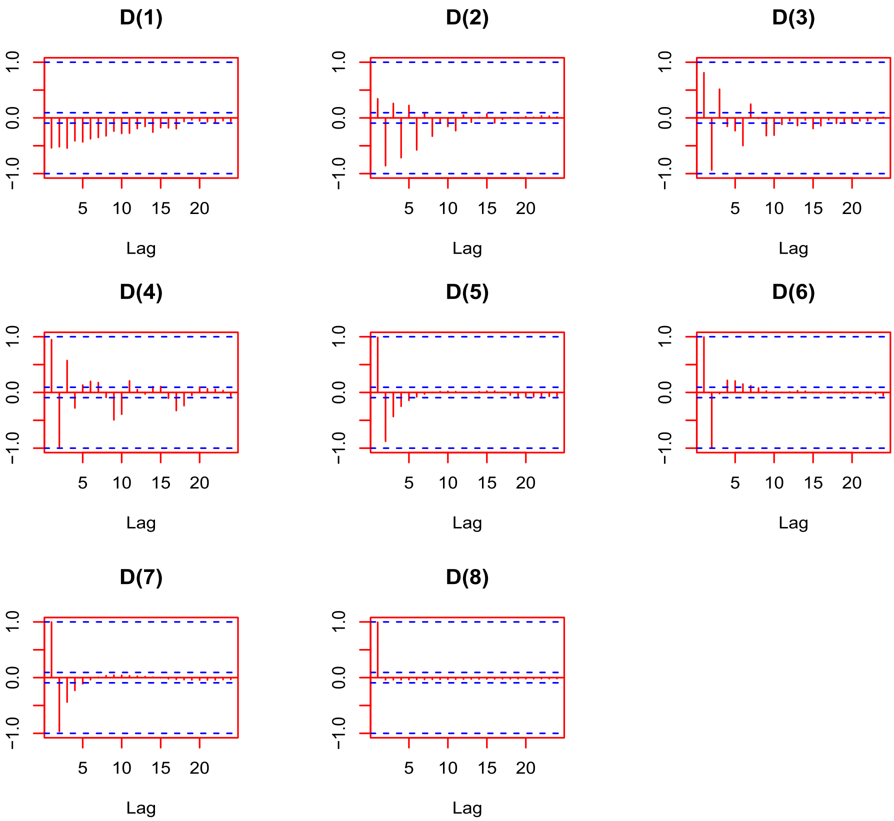


(b) Details

Fig. 9 Wavelet Simple autocorrelation function - LogSoy



(a) Approximation



(b) Details

Fig. 10 Wavelet Partial autocorrelation function - LogSoy

Table 11 Diagnostic of residuals on originals data series models

		Model	Aic	Kpss Test	Ljung-Box Test	Rmse	Mae
Wheat	D(1)	Sarima(3,0,2)(0,0,0)[0]	1718.02	0.10	0.94	1.62	1.05
	D(2)	Sarima(3,0,2)(0,0,0)[0]	1298.81	0.10	0.09	1.01	0.66
	D(3)	Sarima(3,0,1)(0,0,0)[0]	685.98	0.10	0.88	0.51	0.32
	D(4)	Sarima(4,0,1)(0,1,0)[12]	– 772.18	0.10	0.96	0.10	0.07
	D(5)	Sarima(5,0,0)(0,0,1)[0]	– 2184.58	0.10	0.67	0.02	0.01
	D(6)	Sarima(1,0,3)(0,0,1)[0]	– 337.96	0.10	0.00	0.16	0.13
	D(7)	Sarima(2,2,0)(0,0,0)[0]	– 5644.71	0.10	0.00	0.00	0.00
	D(8)	Sarima(1,2,0)(0,0,0)[0]	– 5987.96	0.10	0.00	0.00	0.00
	A(8)	Sarima(2,1,3)(0,0,0)[0]	– 6534.76	0.10	0.00	0.00	0.00
Soy	D(1)	Sarima(3,0,2)(0,0,0)[0]	2760.75	0.10	0.51	5.27	3.50
	D(2)	Sarima(5,0,4)(0,0,0)[0]	1741.47	0.10	0.25	1.61	1.10
	D(3)	Sarima(5,0,2)(0,0,0)[0]	1552.64	0.10	0.65	1.33	0.92
	D(4)	Sarima(3,0,2)(0,1,0)[12]	200.55	0.10	0.27	0.29	0.21
	D(5)	Sarima(4,0,2)(0,0,1)[0]	– 1351.19	0.10	0.97	0.05	0.04
	D(6)	Sarima(1,0,3)(0,0,1)[0]	-1556.37	0.10	0.00	0.04	0.03
	D(7)	Sarima(3,2,0)(0,0,0)[0]	– 4662.31	0.10	0.00	0.00	0.00
	D(8)	Sarima(1,2,0)(0,0,0)[0]	– 4786.38	0.10	0.01	0.00	0.00
	A(8)	Sarima(2,2,3)(0,0,0)[0]	– 6734.74	0.10	0.00	0.00	0.00

for the study of electricity prices, by Rivas et al. (2013) in the detection of cyclical behaviour of metal indices prices. But the gain in this quest for precision with this methodology is to achieve the best configuration that is in the multiple resolution analysis. The hardest part was choosing the wavelet name. So why, in the multi-resolution analysis, we choose other wavelet function in another wavelet family.

These families differ on four main criteria. The length of the support or the compact nature of their support. They have the faculty to represent effectively signals that possess disruptions, discontinuities or abrupt escalations. These characteristics are essential precisely in the share prices of raw materials. In addition, there's the symmetry of forms of wavelets and their number of vanishing moments. Because, the more null moments in the wavelet function the more the transition between the space is smooth. Finally, there is the regularity. It is strongly related to the number of null moments (Tables 12, 13). The Daubechies wavelets are the most completely family. But, the choice of the function depend on the characteristics of times series analysed (Gencay et al., 2001).

This study demonstrates that the combination between the wavelet transform and the seasonal auto-regressive and moving average is of technical interest in the forecast of data series with seasonal components. The commodity price indices like cereals and oleaginous product help to prove how to obtain the best accuracy in

Table 12 Diagnostic of residuals on transforms data series models

		Model	Aic	Kpss Test	Ljung-Box Test	Rmse	Mae
LogWheat	D(1)	Sarima(0,0,3)(0,0,0)[0]	1801.94	0.10	0.56	1.79	1.12
	D(2)	Sarima(5,0,5)(0,0,0)[0]	− 5488.33	0.10	0.08	0.00	0.00
	D(3)	Sarima(5,0,5)(0,0,0)[0]	− 6014.32	0.10	0.66	0.00	0.00
	D(4)	Sarima(3,0,0)(0,1,0)[12]	− 6522.19	0.10	0.97	0.00	0.00
	D(5)	Sarima(5,0,0)(0,0,0)[0]	− 8508.98	0.10	0.93	0.00	0.00
	D(6)	Sarima(1,0,3)(0,0,1)[0]	− 10375.2	0.10	0.53	0.00	0.00
	D(7)	Sarima(2,2,0)(0,0,0)[0]	− 11974.3	0.10	0.16	0.00	0.00
	D(8)	Sarima(1,2,0)(0,0,0)[0]	− 13612.6	0.10	0.31	0.00	0.00
	A(8)	Sarima(2,1,3)(0,0,0)[0]	− 10544.5	0.10	0.00	0.00	0.00
LogSoy	D(1)	Sarima(3,0,2)(0,0,0)[0]	− 4572.48	0.10	0.04	0.00	0.00
	D(2)	Sarima(5,0,4)(0,0,0)[0]	− 5990.13	0.10	0.21	0.00	0.00
	D(3)	Sarima(5,0,2)(0,0,0)[0]	− 5828.33	0.10	0.07	0.00	0.00
	D(4)	Sarima(3,0,2)(0,1,0)[12]	− 7246.70	0.10	0.69	0.00	0.00
	D(5)	Sarima(4,0,2)(0,0,1)[0]	− 8740.81	0.10	0.35	0.00	0.00
	D(6)	Sarima(1,0,3)(0,0,1)[0]	− 9529.07	0.10	0.00	0.00	0.00
	D(7)	Sarima(3,2,0)(0,0,0)[0]	− 12094.69	0.10	0.39	0.00	0.00
	D(8)	Sarima(1,2,0)(0,0,0)[0]	− 13826.1	0.10	0.01	0.00	0.00
	A(8)	Sarima(2,2,3)(0,0,0)[0]	− -15754.0	0.10	0.99	0.00	0.00

Table 13 Forecast and Comparison

		Wavelet-Sarima		Sarima	
	Wavelet	Rmse	Mae	Rmse	Mae
Wheat	Modwt-d4	0.80	0.34	12.05	7.76
LogWheat	Modwt-d4	0.00	0.00		
Soy	Modwt-d4	3.60	1.56	36.95	25.27
LogSoy	Modwt-d4	0.00	0.00		

forecast. But, it is necessary to have the optimal setting in multi-resolution analysis. In this configuration the main parameter is the choice of wavelet function. By referring to the previous research done in economic and financial data series, the Daubechies's wavelet is the most used. So, the same support has been used in these multi-scales applications with agricultural commodities. But, according to the type of data analysis and the object of the study, this resolution can be made by using another type of wavelet.

Appendix

See Figs. 11, 12, 13, 14, 15, 16, 17, 18, 19, 20.

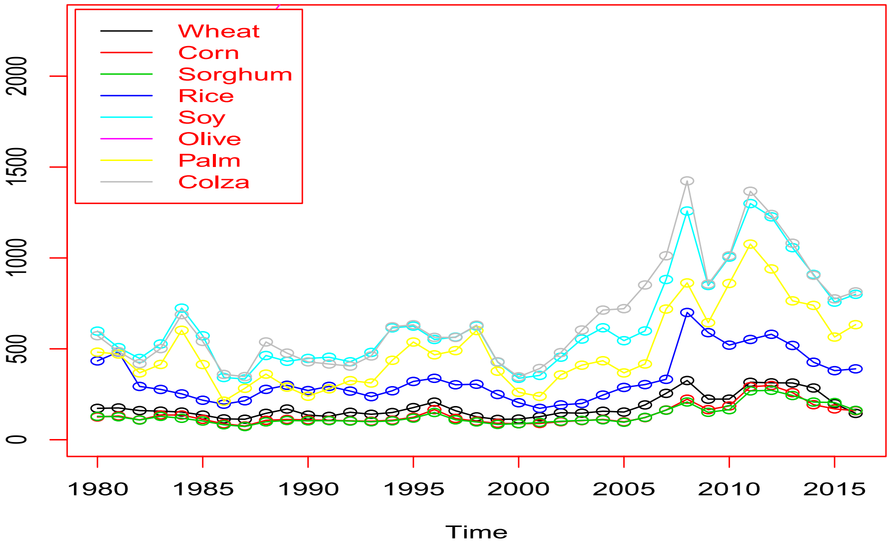


Fig. 11 Annual average data series

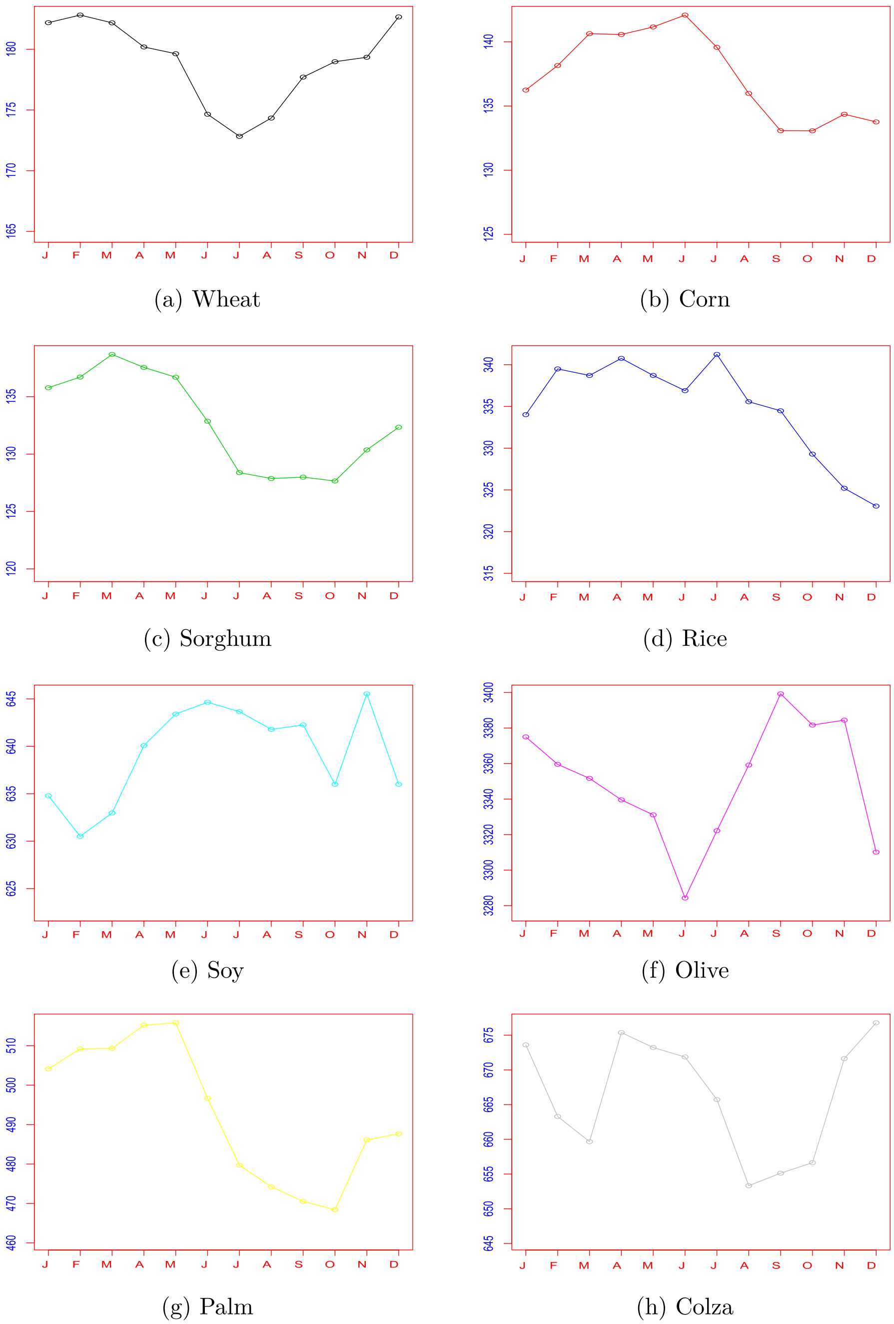
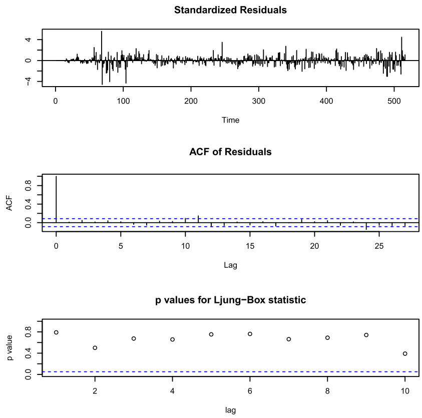
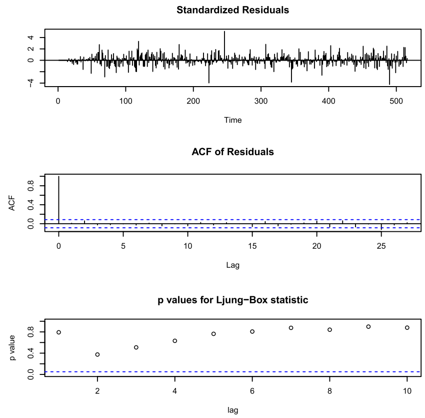


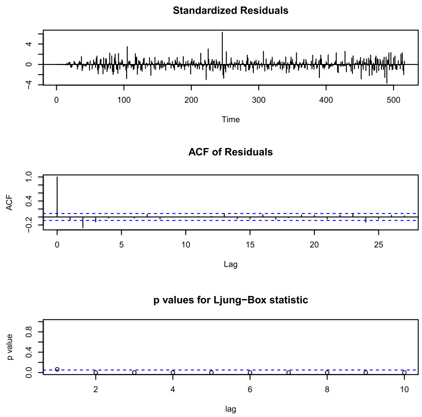
Fig. 12 Monthly average indices prices



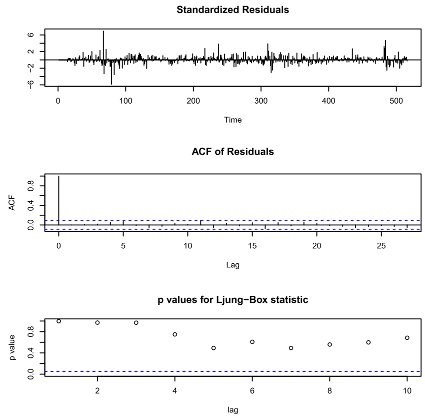
(a) Wheat



(b) Corn

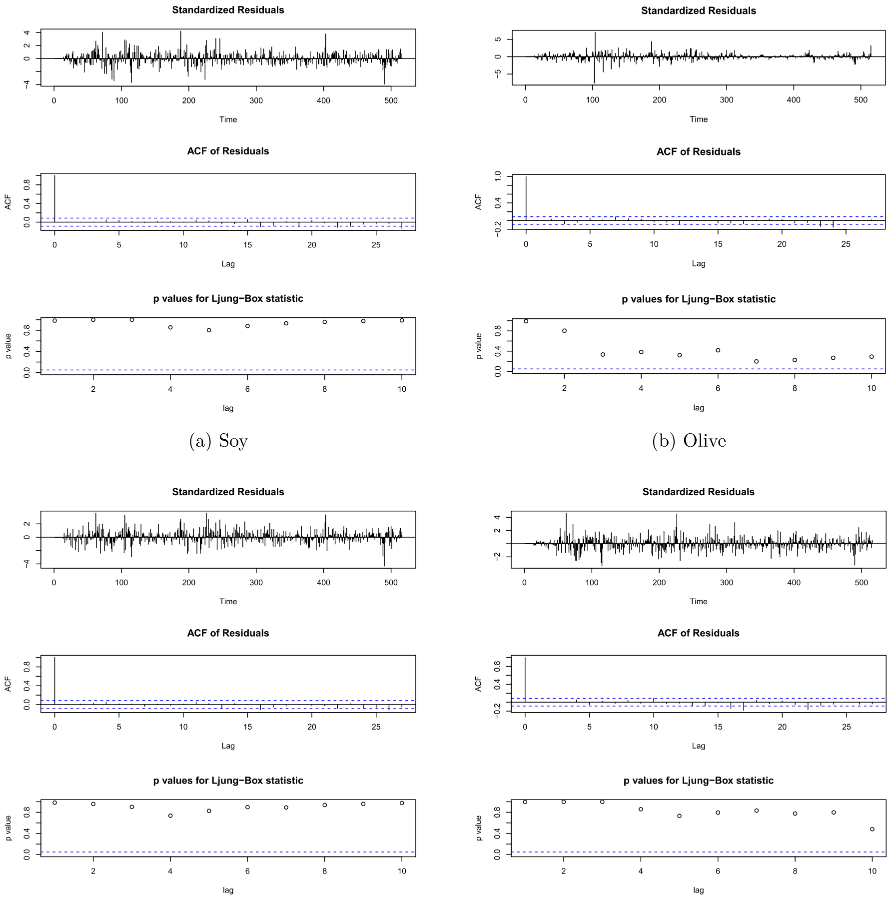


(c) Sorghum



(d) Rice

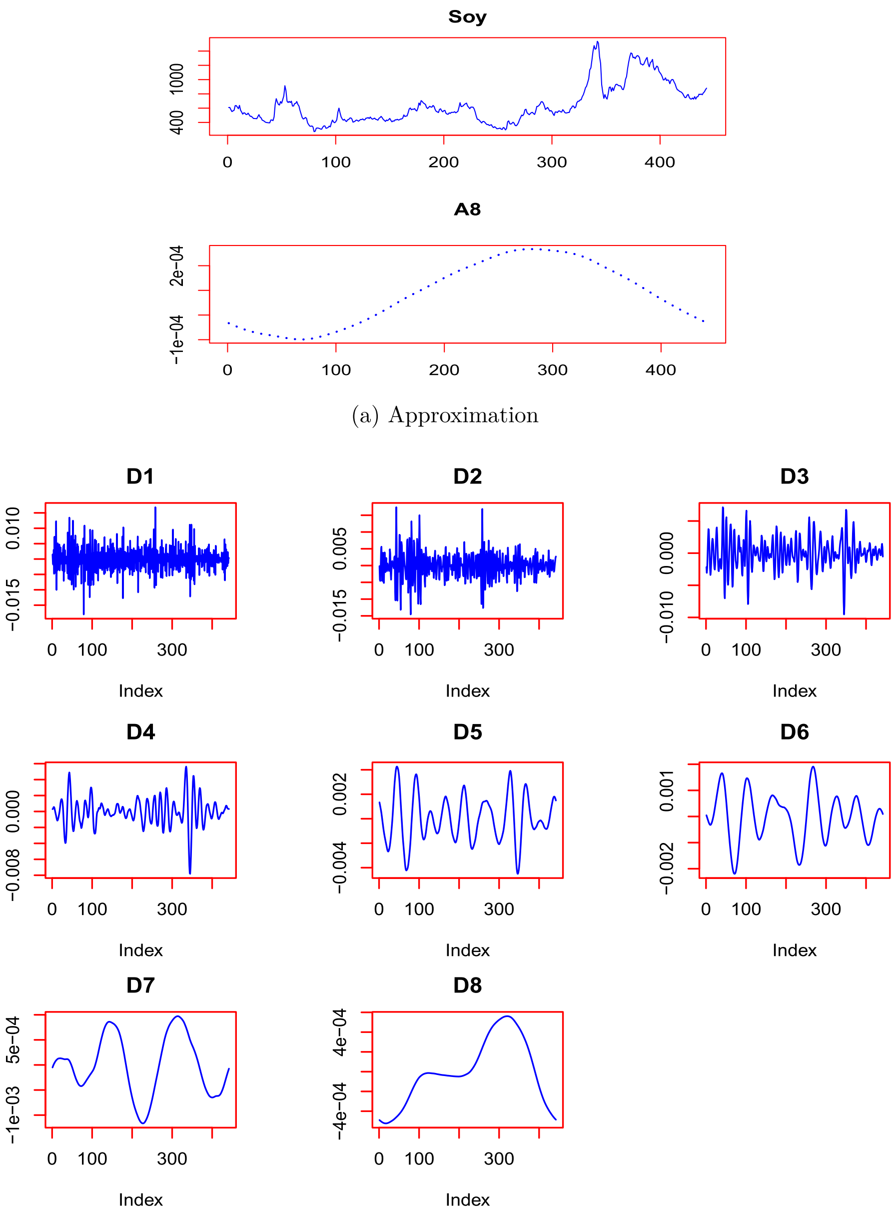
Fig. 13 Residuals diagnostics by Seasonal Arima model for cereals data series



(a) Soy (b) Olive

(c) Palm (d) Colza

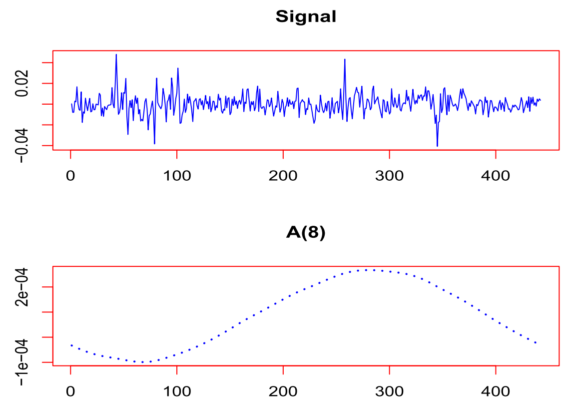
Fig. 14 Residuals diagnostics by Seasonal Arima model for oleaginous data series



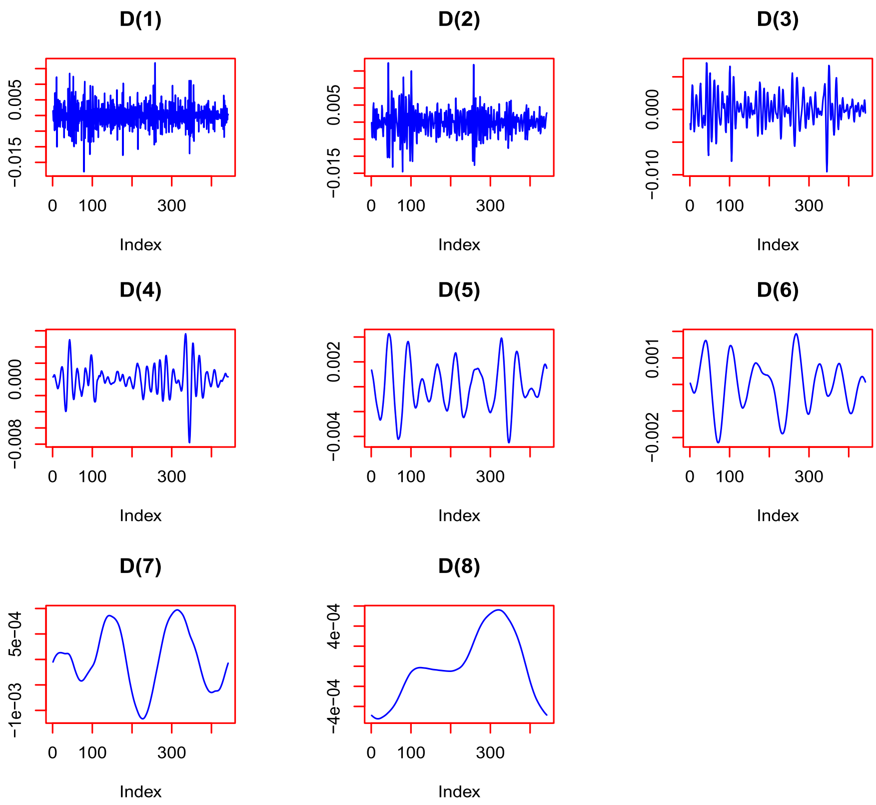
(a) Approximation

(b) Details

Fig. 15 MRA on Soy times series

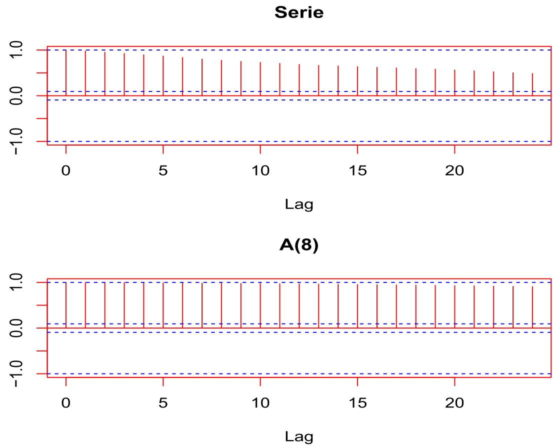


(a) Approximation

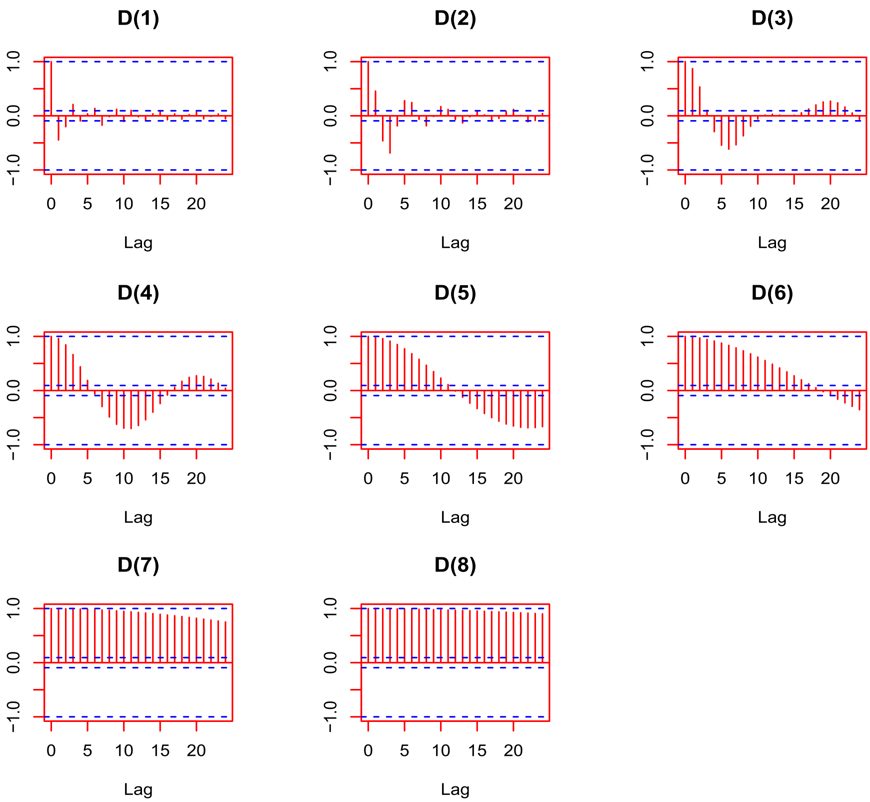


(b) Details

Fig. 16 MRA on LogSoy times series

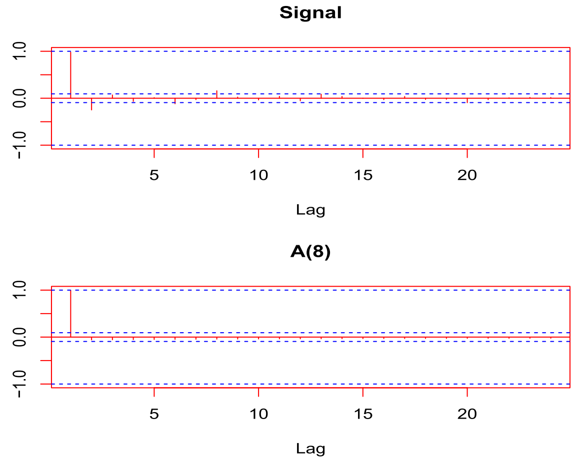


(a) Approximation

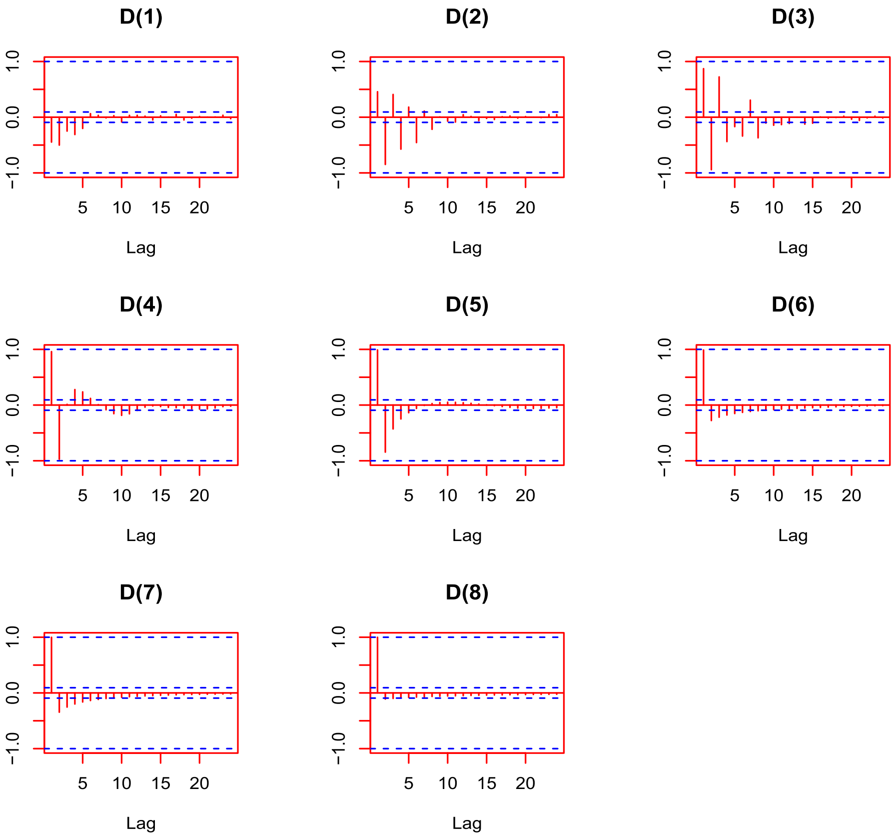


(b) Details

Fig. 17 Wavelet Simple autocorrelation function - Wheat

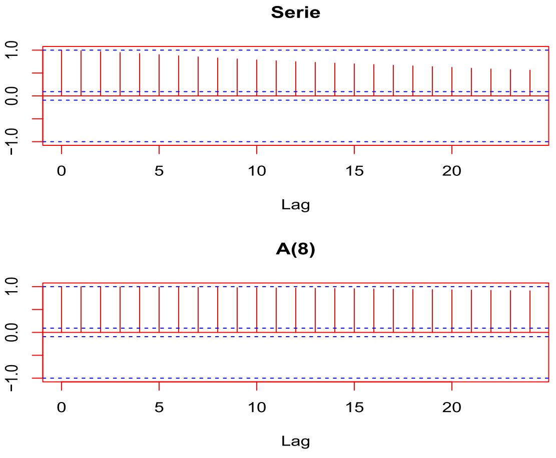


(a) Approximation

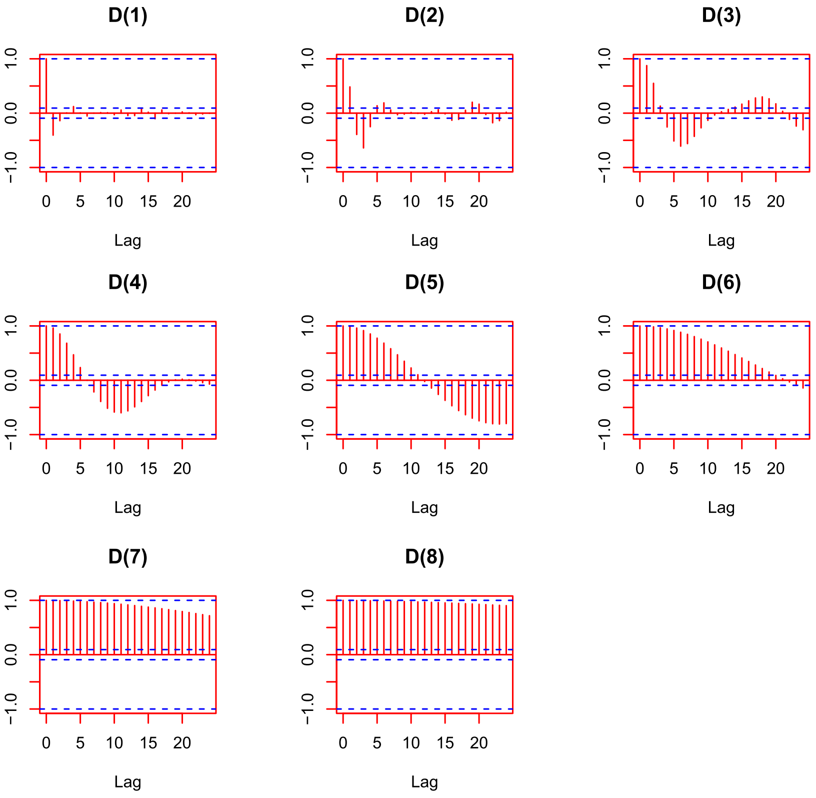


(b) Details

Fig. 18 Wavelet Partial autocorrelation function - Wheat

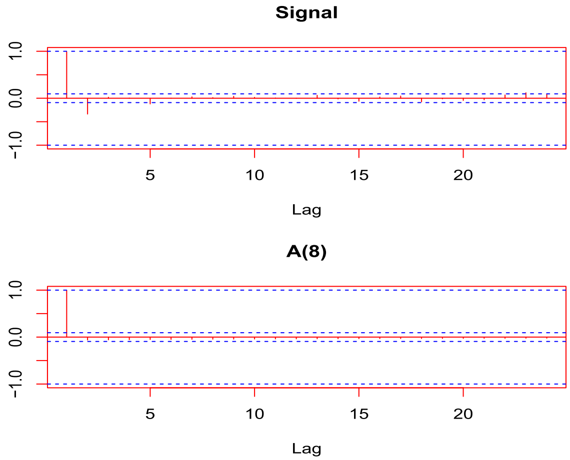


(a) Approximation

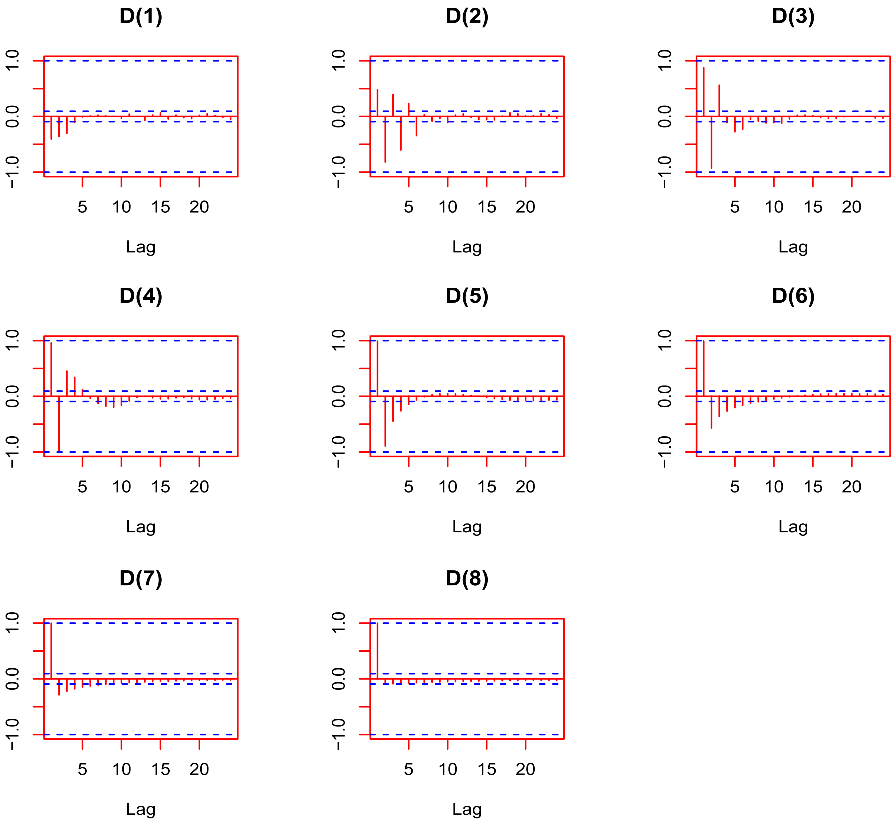


(b) Details

Fig. 19 Wavelet Simple autocorrelation function - Soy



(a) Approximation



(b) Details

Fig. 20 Wavelet Partial autocorrelation function - Soy

References

Abry, P. 1997. *Ondelettes et turbulences* Nouveaux essais , arts et sciences, Diderot.

Box, G.E., and D. R. Cox. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society*. Series B(Methodological), 26: 211–252

Box, G.E., and G.M. Jenkins. (1976) Time series analysis forecasting and control. *Rev*.

Box, G.E., G.M. Jenkins, G.C. Reinsel, and G.M. Ljung. 2015. *Time series analysis: Forecasting and control*. London: John Wiley Sons.

Chao, S., and Y.C. He. 2015. SVM-ARIMA agricultural product price forecasting model based on wavelet decomposition. *Statistics and Decision* 13: 92–95.

Choudhary, K., G.K. Jha, R.R. Kumar, and D.C. Mishra. 2019. Agricultural commodity price analysis using ensemble empirical mode decomposition: A case study of daily potato price series. *Indian Journal of Agricultural Sciences* 89 (5): 882–886.

Cleveland, R.B., W.S. Cleveland, J.E. McRae, and I. Terpenning. 1990. STL: A seasonal trend decomposition procedure based on loss. *Journal of Official Statistics* 6 (1): 3–73.

Conejo, A.J., J. Contreras, R. Espínola, et al. (2005). Forecasting electricity prices for a day-ahead pool-based electric energy market. *International journal of forecasting* 21(3):435–462

Daubechies, I. 1992. Ten lectures on wavelets. SIAM.

Dickey, D.A.., W.A. Fuller. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association* 74: 427–431.

Gomez, V., and A. Maravall. (1998). Seasonal adjustment and signal extraction in economic time series. Documentos de trabajo / Banco de España 9809, 21-abr-1998, ISBN: 847793598X

Gencay, R., F. Selçuk, and B. Whitcher. 2001. *An introduction to wavelets and other filtering methods in finance and economics*. San Diego: Academic Press.

Gencay, R., F. Selçuk, and B. Whitcher. 2005. Multiscale systematic risk. *Journal of International Money and Finance* 24 (1): 55–70.

Ghysels, E. 1998. On stable factor structures in the pricing of risk: Do time-varying betas help or hurt? *Journal of Finance* 53 (2): 549–573.

Hannan, E.J., J. Rissanen. (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika* 69(1):81–94

Hayat, A., and M.I. Bhatti. 2013. Masking of volatility by seasonal adjustment methods. *Economic Modelling* 33: 676–688. https://doi.org/10.1016/j.econmod.2013.05.016.

Hyndman, R. J., and Y. Khandakar. (2008). Automatic Time Series Forecasting: The forecast Package for R. Journal of Statistical Software 27(3):1–22

Hyndman, R., A. Koehler, K. Ord, et al. (2008). Forecasting with exponential smoothing: the state space approach. Springer Science & Business Media.

Jadhav, V., B.V.C. Reddy, and G.M. Gaddi. 2018. Application of ARIMA model for forecasting agricultural prices. *Journal of Agriculture Science and Technology A* 19 (5): 981–992.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin. (1994). Testing the Null Hypothesis ofStationarity against the Alternative of a Unit Root. Journal of Econometrics. Vol. 54, pp. 159-178. 2. Hamilton, J. D. Time Series Analysis. Princeton, NJ: Princeton University Press.

Levhari, D., and H. Levy. 1977. The capital asset pricing model and the investment horizon. *The Review of Economics and Statistics* 59 (1): 92–104.

Li, B., J. Ding, Z. Yin, K. Li, X. Zhao, and L. Zhang. 2021. Optimized neural network combined model based on the induced ordered weighted averaging operator for vegetable price forecasting. *Expert Systems with Applications* 168: 114–232. https://doi.org/10.1016/j.eswa.2020.114232.

Liu, C.-Y. and Z.-Y. Zheng. (1989). Stabilization Coefficient of Random Variable. Biom. J. 31: 431–441.

Mallat, Stéphane. 1989. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11 (7): 674–693.

Melard, G., and J.M., Pasteels. (2000). Automatic ARIMA modeling including interventions, using time series expert software. *International Journal of Forecasting* 16(4): 497–508

Misiti, M., Y. Misiti, G. Oppenheim, and J. M. Poggi. (2003). Les ondelettes et leurs applications. Hermès science publications.

OCDE. (2008). Rapport annuel de l'OCDE 2008, Éditions OCDE, Paris. https://doi.org/10.1787/annrep-2008-fr

Osborn D.R., A.P.L. Chui, P.J.P. Smith, C.R. Birchenhall. (1988). Seasonality and the order of integration for consumption. *Oxford Bulletin of Economics and Statistics* 50(4): 361–77

Philips, P.C.B., P. Perron. (1987). Testing for a Unit Root in Time Series Regression. *Biometrika* 75:335–346.

Rivas, M. et al. (2013). Linking the energy system and ecosystem services in real landscapes. *Biomass and Bioenergy* 55:17–26

Sadefo Kamdem, J., A. Nsouadi, and M. Terraza. 2016. Time-frequency analysis of the relationship between EUA and CER carbon markets. *Environmental Modeling and Assessment* 21: 279–289.

Shumway, R.H., D.S. Stoffer. (2006). Time series regression and exploratory data analysis. *Time Series Analysis and Its Applications: With R Examples* 48–83

Unser, M. 1996. *Wavelet in medecine and biology*. London: CRC Press.

Vannucci, M., and F. Corradi. 1999. Covariance structure of wavelet coefficients: Theory and models in a Bayesian perspective. *Journal of Royal Statistical Society B* 4: 971–986.

Wang, J., Z. Wang, X. Li, and H. Zhou. 2019. Artificial bee colony-based combination approach to forecasting agricultural commodity prices. *International Journal of Forecasting*. https://doi.org/10.1016/j.ijforecast.2019.08.006.

Xiong, T., C. Li, and Y. Bao. 2018. Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China. *Neurocomputing* 275: 2831–2844. https://doi.org/10.1016/j.neucom.2017.11.053.