



# O zastosowaniu metody widma osobliwego do analizy serii czasowych wielkości hydrofizycznych

(poziom morza na wybranych stacjach mareograficznych w rejonie Morza Północnego i Bałtyckiego)

## Cześć pierwsza - part 1

( Na bazie prezentacji na Seminarium Zakładu Dynamiki Morza - 08.12.2015 r.) <sup>a</sup>

Andrzej Jankowski

Institute of Oceanology of PAS, Powstancow Warszawy 55, 81-712 Sopot, Poland

e-mail: jankowsk@iopan.gda.pl

---

<sup>a</sup>© Andrzej Jankowski, 2015

## Spis treści wystąpienia

- Wstęp
- Metoda ( Zarys metody)
- Serie syntetyczne - modelowe
- Zakres i wyniki prac obliczeniowych
- Uwagi końcowe

# Cześć pierwsza - part 1

## Wstęp

Metoda widma osobliwego (Singular Spectrum Analysis - SSA) - to jedna z metod zalecanych i stosowanych do analizy klimatycznych serii czasowych ( np. Ghil i in., 2002 <sup>a</sup>; Broomhead i King, 1986 <sup>b</sup>; Vautard i Ghil (1989) <sup>c</sup>; Różyński, 2003 <sup>d</sup>; Pruszek i in., 2004 <sup>e</sup>; Kaczmarek i in., 2005 <sup>f</sup> ).

Metoda umożliwia identyfikację składowych procesów formujących zmienność wybranej wielkości z analizy spektralnej zaszu- mionej serii pomiarowej: wydzielenie trendu (+ b. wolnozmiennie składowe), składowych oscylacyjnych, składowych losowych oraz składowych pozbawionych określonej struktury czyli szumów.

Różyński (2003) pokazał zalety metody widma osobliwego do identyfikacji wymuszonych i chaotycznych elementów ewolucji linii brzegowej w rejonie Lubiatowa.

Pruszek i in. (2004) wykorzystali, z powodzeniem, metodę analizy widma osobliwego (SSA) do analizy wyników pomiarów terenowych fal pogravitacyjnych co umożliwiło identyfikację progresywnych fal krańcowych i odpowiadających im liczb modal- nych, jakie mogą występować w warunkach wielorewowej strefy południowego Bałtyku.

Kaczmarek i in. (2005) pokazali przydatność metod SSA w badaniach procesów hydrodynamicznych w strefie przybrzeżnej. Technika zastosowana do serii pomiarów prądów falowych wzdłużbrzegowych potwierdziła wysoką precyzję metody w wyodrębnianiu (wydzieleniu) składowych oscylacyjnych dla określonych, zawężonych, przedziałów częstotliwości.

## Cel pracy

- zastosowanie metody do analizy zmienności poziomu morza z pomiarów na wybranych stacjach mereograficznych

---

<sup>a</sup>Ghil M., R. M. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, et al. (2002), *Advanced spectral methods for climatic time series*, Rev. Geophys. 40(1), 3.1-3.41, doi: 10.1029/2000RG000092

<sup>b</sup>Broomhead, D. S. and King, G. P. (1986). Extracting qualitative dynamics from experimental data. *Physica D* 20, 217-236

<sup>c</sup>Vautard, R., Ghil, M., 1989, *Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series*, *Physica D*, 35, 395-424.

<sup>d</sup>Różyński G., (2003), *Coastal nearshore morphology in terms of large data sets*, IBW PAN, Gdańsk, 1-170

<sup>e</sup>Pruszek Z. Różyński G. Szmytkiewicz M., 2004, *Fale pogravitacyjne w warunkach południowego Bałtyku*, *Inżynieria Morska i Geotechnika*, nr 1, s. 8 - 15

<sup>f</sup>Kaczmarek J., Różyński G., Pruszek Z., 2005, *Long period oscillations in the longshore current on a sandy, barred coast investigated with Singular Spectrum Analysis*, *Oceanologia* 2005, no 47(1), pp. 5-25

## Metoda (Zarys metody)

Szersze i bardziej szczegółowe omówienie metodologii - można znaleźć w pracach (Vautard i Ghil, 1989; Hassani, 2007<sup>a</sup>; Golyandina i in., 2001<sup>b</sup>; Golyandina i Zhigljavsky, 2013<sup>c</sup>; Elsner i Tsonis, 1996<sup>d</sup>).

W praktycznych aplikacjach, metoda SSA, stanowi nieparametryczną metodę estymacji spektralnej wyjściowej serii czasowej, której stosowanie można przedstawić w następujących etapach (krokach).

etap 1 - dekompozycja -rozłożenie wyjściowej serii czasowej:

$$y(t); t = 1, \dots, N$$

$$\{y(t) :, t = 1, \dots, N\} \longrightarrow \{y(t - j) :, t = 1, \dots, N; j = 1, L\} \longrightarrow \{y(k) :, k = 1, \dots, N - L + 1\}$$

na elementy  $L$ -wymiarowej przestrzeni wektorowej  $L$  ( $1 < L < N$ )

poprzez konstrukcję specyficznej macierzy o wymiarze  $L \times K$ , gdzie  $K = N - L + 1$ , zwanej macierzą *trajektorii*:

$$\mathbf{X} = \begin{bmatrix} y(1) & y(2) & y(3) & \dots & y(K) \\ y(2) & y(3) & y(4) & \dots & y(K + 1) \\ y(3) & y(4) & y(5) & \dots & y(K + 2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y(L) & y(L + 1) & y(L + 2) & \dots & y(N) \end{bmatrix};$$

<sup>a</sup>Hassani H., 2007, *Singular Spectrum Analysis: Methodology and Comparison*, Journal of Data Science 5. (2007): 239-257.

<sup>b</sup>Golyandina, N., Nekrutkin, V., Zhigljavsky, A., (2001), *Analysis of Time Series Structure: SSA and Related Techniques*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability (Book 90), pp. 320

<sup>c</sup>Golyandina N., Zhigljavsky A., 2013, *Singular Spectrum Analysis for Time Series*, SpringerBriefs in Statistics, Springer, pp. 120

<sup>d</sup>Elsner J.B., Tsonis A.A., (1996), *Singular Spectrum Analysis. A New Tool in Time Series Analysis*, Plenum Press, New York

etap 2 - wyznaczenie  $L \times L$  macierzy kowariancji:

$$\mathbf{C}_X = \frac{1}{N} \mathbf{X} \mathbf{X}^T$$

gdzie Symbol  $T$  - oznacza operację transpozycji macierzy lub wektora.

etap 3 - diagoanlizacja macierzy kowariacji:

- wyznaczenie wartości własnych i odpowiadających im wektorów własnych macierzy

$$\mathbf{E}^T \mathbf{C}_X \mathbf{E} = \mathbf{\Lambda} \quad \longrightarrow \quad \mathbf{E}^T \mathbf{X} \mathbf{X}^T \mathbf{E} = \mathbf{\Lambda} N$$

gdzie:

$\mathbf{\Lambda}$  - macierz diagonalna ( $L \times L$ ), zawierająca wartości własne  $\{\lambda_k; 1 \leq k \leq L\}$

$\mathbf{E}$  - macierz ( $L \times L$ ), której kolumny zawierają wektory własne  $\{\mathbf{E}_k; 1 \leq k \leq L\}$

Wartości własne  $\lambda_k$ ,  $k = 1, \dots, L$  wyrażają wariancję analizowanej serii czasowej w kierunku wskazanym przez odpowiednie wektory własne  $\mathbf{E}_k$ ,  $k = 1, \dots, L$ .

Pierwiastki kwadratowe z wartości własnych to - tzw. wartości osobliwe (singularne), a ciąg wartości  $\{\sqrt{\lambda_k}; 1 \leq k \leq L\}$  - tzw. widmo osobliwe (singularne).

etap 4 - wyznaczenie składowych głównych (Principal components)

Projekcja wyjściowej serii czasowej  $y(t)$  na poszczególne wektory własne  $\mathbf{E}_k$  o składowych  $\{E_{k,j}; j = 1, \dots, L\}$

daje oszacowanie odpowiadających im czasowych składowych głównych (Principal Components - PCs) -  $A_k(t); k = 1, \dots, L$ :

$$A_k(t) = \sum_{j=1}^L y(t+j-1) E_{k,j}; \quad t = 1, \dots, N-L+1; \quad k = 1, \dots, L$$

etap 5 - wydzielenie struktur - analiza widma wartości własnych  $\lambda_k$ ,  $k = 1, \dots, L$

Wartości własne  $\lambda_k$ ,  $k = 1, \dots, L$  wyrażają variancję analizowanej serii czasowej w kierunku wskazanym przez odpowiednie wektory własne  $\mathbf{E}_k$ ,  $k = 1, \dots, L$ .

Wartości własne  $\lambda_k$ ;  $k = 1, \dots, L$  macierzy trajektorii wyrażają wartość, oszacowanie variancji cząstkowej dla  $k$ -tego wektora własnego, a ich suma wyraża variancję całkowitą analizowanej wyjściowej serii czasowej:

$$W_c = \sum_{i=1}^L \lambda_i = \sum_{i=1}^L W_i; \quad \Leftrightarrow \quad W_i = \lambda_i; \quad i = 1, \dots, L$$

Zatem tzw. variancję wyjaśnioną dla danego elementu (składnika) szeregu czasowego (serii) możemy zapisać jako:

$$W_m^{exp} = \frac{W_m}{W_c} = \frac{\lambda_m}{W_c} = \frac{\lambda_m}{\sum_{i=1}^L \lambda_i}; \quad m = 1, \dots, L$$

Jeśli uwzględnimy wszystkie składowe to otrzymamy:

$$\sum_{m=1}^L W_m^{exp} = \sum_{m=1}^L \frac{W_m}{W_c} = \sum_{m=1}^L \frac{\lambda_m}{W_c} = \frac{\sum_{m=1}^L \lambda_m}{\sum_{i=1}^L \lambda_i} = 1$$

Zatem, serię wyjściową, rozłożono na sumę składowych (modów, struktur), wśród których można wydzielić te, które reprezentują trend, oscylacje czy szum.

Można dokonać tego analizując graficzną postać widma wartości własnych  $\lambda_k$  dla kolejnych wartości indeksu  $k$  - tzw. **scree diagram**.

Punkt przełamania na wykresie daje zgrubne oszacowanie pożytecznego sygnału, którego wkład w pełną variancję sygnału (objaśnianą) jest maksymalny dla zadanego poziomu dokładności wkładu variancji cząstkowych w objaśnianą - całkowitą serię wyjściową.

etap 6 - rekonstrukcja - powrót serii wyjściowej

Sumując, bez wydzielenia grup, mamy (pojedyncze struktury, mody):

$$R_k(t) = \frac{1}{L_t} \sum_{j=L_t}^L A_k(t-j)E_{k,j}; \quad 1 \leq t < L; \quad L_t = \min(t, N-t+1); \quad 1 \leq j \leq t$$

$$R_k(t) = \frac{1}{L_t} \sum_{j=L_t}^L A_k(t-j)E_{k,j}; \quad L \leq t \leq K; \quad L_t = L; \quad t \leq j \leq t-N+L-1$$

$$R_k(t) = \frac{1}{L_t} \sum_{j=L_t}^L A_k(t-j)E_{k,j}; \quad K+1 \leq t \leq N; \quad L_t = \min(t, N-t+1); \quad t-N+L \leq j \leq L$$

Sumując, otrzymamy, zrekonstruowaną serię wyjściową:

$$\tilde{y}(t) = \sum_{k=1}^L R_k(t); \quad t = 1, \dots, N$$

## Serie syntetyczne - modelowe

Seria czasowa  $y(t)$ ;  $t = 1, \dots, N$  składająca się z sygnału harmonicznego (cykl roczny - 6 składowych harmonicznnych o okresach:  $T=12$ mies., 6 mies, 4, 3, 2.4 i 2 mies) oraz dodatkowego członu (szum biały):

$$X(t) = \sum_{i=1}^6 C_i \left( A_i \cos\left(\frac{2\pi}{T} it\right) + B_i \sin\left(\frac{2\pi}{T} it\right) \right) + \zeta(t); \quad t = 1, \dots, N$$

gdzie:  $\zeta(t)$  - szum;  $T_o = 12$ mies.- okres rocznej harmoniki,  $N$ - liczba elementów szeregu czasowego;  $i = 1, 2, \dots, 6$  - numer kolejnej harmoniki;  $t$ - czas (w mies.),  $C_i$  - współczynniki korygujące wartość amplitudy składowej cyklu rocznego,  $\zeta(t)$  - człon opisujący szum biały:

$$\zeta(t) = coef \times \epsilon(t); \quad t = 1, \dots, N;$$

gdzie:  $\epsilon(t)$  - szum - seria czasowa o parametrach (wartości średniej i wariancji):  $\bar{\epsilon} = 0$  - wartość średnia;  $\overline{\epsilon'^2} = 1$  - wartość wariancji; wyznaczony z generatora o znormalizowanym rozkładzie normalnym (szum gaussowski).

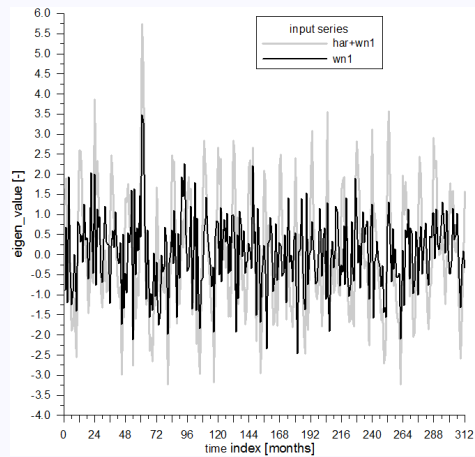
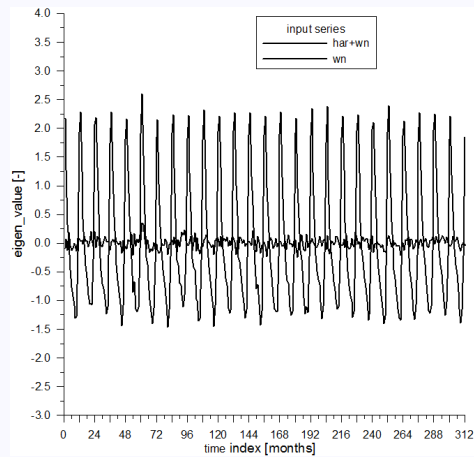
Wartości amplitud  $A_i; B_i$  oraz współczynników je korygujących  $C_i$  prezentuje poniższa tabela.

Model 1 - **har+wn** - sygnał harmoniczny (suma składowych cyklu rocznego) + szum biały - współczynnik -  $coef = 0.1$

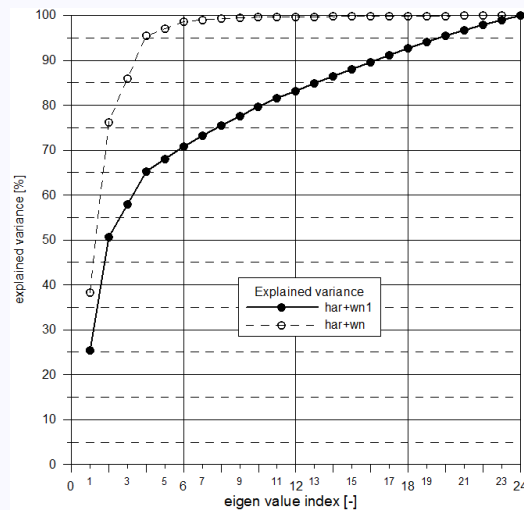
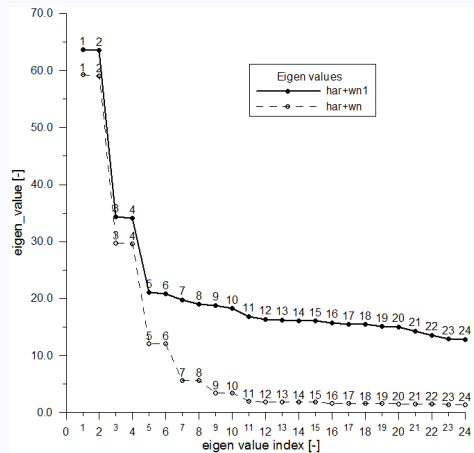
Model 2 - **har+wn1** - sygnał harmoniczny (suma składowych cyklu rocznego) + szum biały - współczynnik -  $coef = 1.0$

No.(i)	Okres	$A_i$	$B_i$	$C_i$
1	12.0	1.0	1.0	1.00
2	6.0	1.0	1.0	0.50
3	4.0	1.0	1.0	0.20
4	3.0	1.0	1.0	0.10
5	2.4	1.0	1.0	0.05
6	2.0	1.0	1.0	0.01

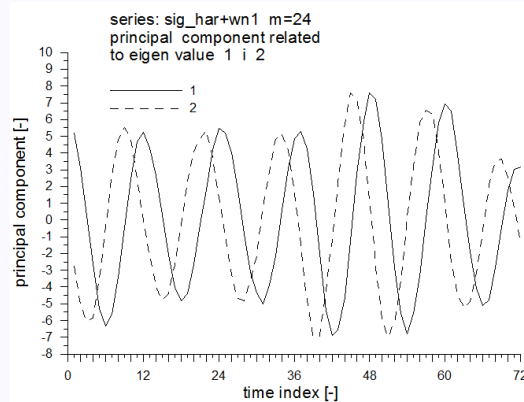
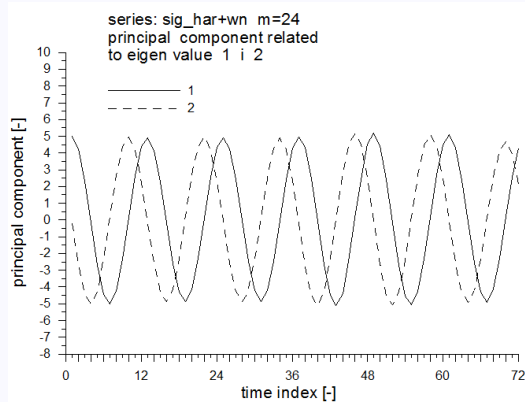




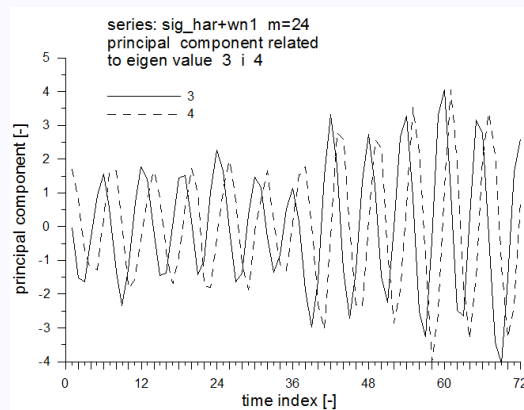
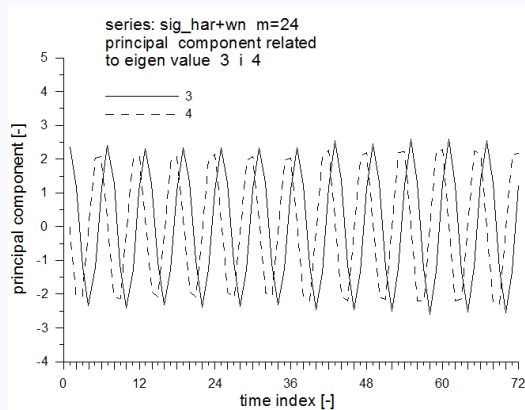
Rys. s1 Serie czasowe wejściowe: model 1 -  $har+wn$  (rysunek lewy); model 2 -  $har+wn1$  (rysunek prawy)



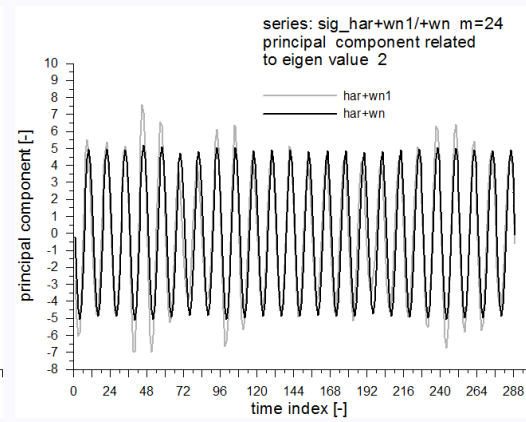
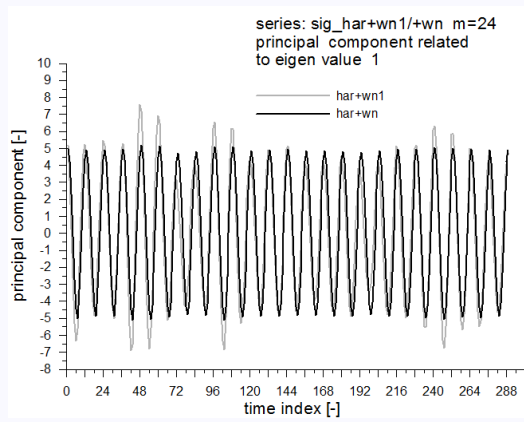
Rys. s2 Widmo wartości własnych (rysunek lewy); wariacja objaśniona (rysunek prawy), dla serii czasowych modelu 1 -  $har+wn$  i modelu 2 -  $har+wn1$ .



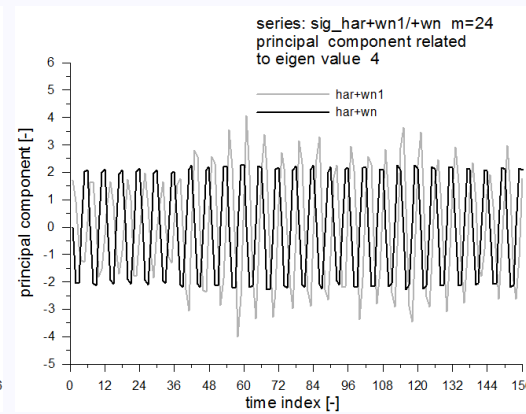
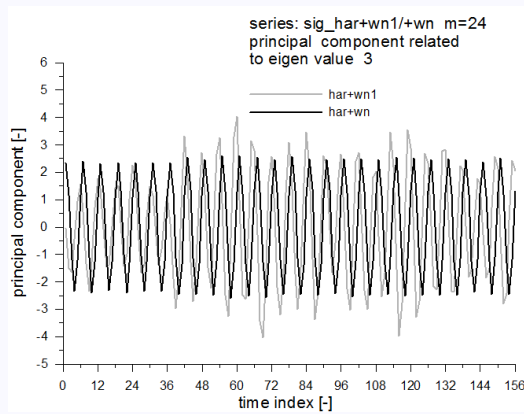
Rys. s3 Składowe główne (principal components) dla wartości własnej 1 i 2 serii czasowych, odpowiednio, modelu 1 -  $\mathbf{har}+\mathbf{wn}$  (rysunek lewy ) oraz modelu 2 -  $\mathbf{har}+\mathbf{wn1}$  (rysunek prawy).



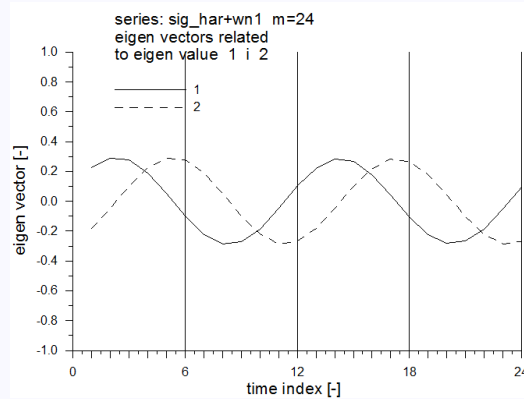
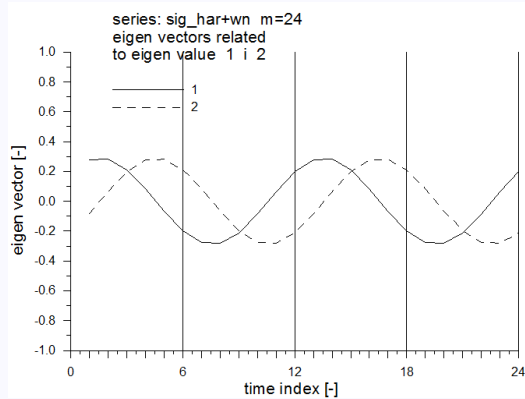
Rys. s4 Składowe główne (principal components) dla wartości własnej 3 i 4 serii czasowych, odpowiednio, modelu 1 -  $\mathbf{har}+\mathbf{wn}$  (rysunek lewy ) oraz modelu 2 -  $\mathbf{har}+\mathbf{wn1}$  (rysunek prawy).



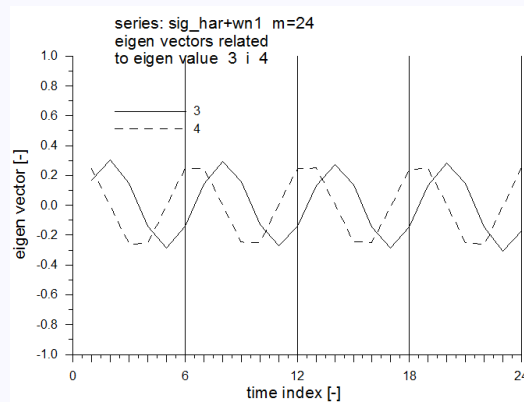
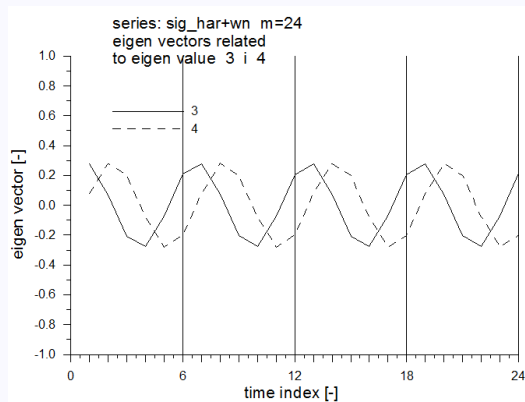
Rys. s5 Składowe główne (principal components), odpowiednio dla wartości własnej 1 (rysunek lewy) oraz 2 (rysunek prawy), dla serii czasowych modelu 1 -  $\mathbf{har} + \mathbf{wn}$  i modelu 2 -  $\mathbf{har} + \mathbf{wn1}$ .



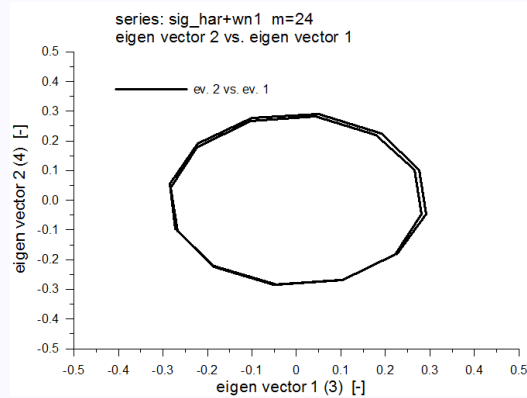
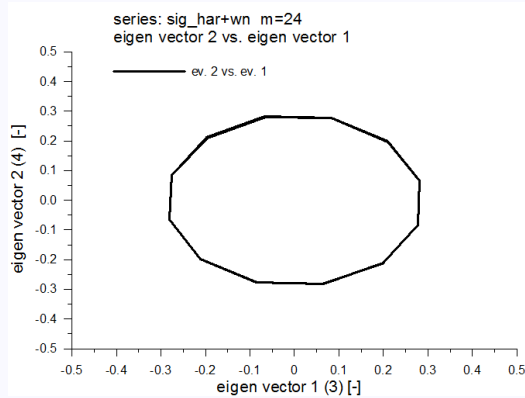
Rys. s6 Składowe główne (principal components), odpowiednio dla wartości własnej 3 (rysunek lewy) oraz 4 (rysunek prawy), dla serii czasowych modelu 1 -  $\mathbf{har} + \mathbf{wn}$  i modelu 2 -  $\mathbf{har} + \mathbf{wn1}$ .



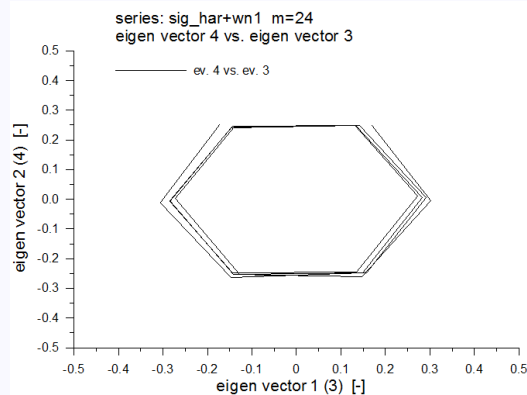
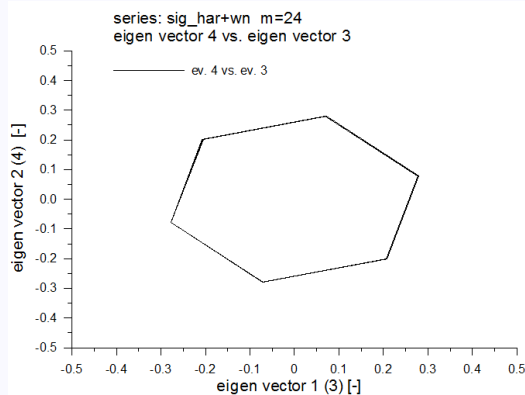
Rys. s7 Wektory własne dla wartości własnej 1 i 2 serii czasowych, odpowiednio, modelu 1 -  $\text{har} + \text{wn}$  (rysunek lewy) oraz modelu 2 -  $\text{har} + \text{wn1}$  (rysunek prawy).



Rys. s8 Wektory własne dla wartości własnej 3 i 4 serii czasowych, odpowiednio, modelu 1 -  $\text{har} + \text{wn}$  (rysunek lewy) oraz modelu 2 -  $\text{har} + \text{wn1}$  (rysunek prawy).



Rys. s9 Wektor własny dla wartości własnej 2 vs. wektor własny dla własnej 1 serii czasowych, odpowiednio, modelu 1 -  $\text{har} + \text{wn}$  (rysunek lewy) oraz modelu 2 -  $\text{har} + \text{wn1}$  (rysunek prawy).



Rys. s10 Wektor własny dla wartości własnej 4 vs. wektor własny dla własnej 3 serii czasowych, odpowiednio, modelu 1 -  $\text{har} + \text{wn}$  (rysunek lewy) oraz modelu 2 -  $\text{har} + \text{wn1}$  (rysunek prawy).1

Cześć pierwsza - part 1 - koniec - end