Time series analysis of complex systems (複雜系統的時間序列分析)

Ming-Chya Wu (吳明佳)

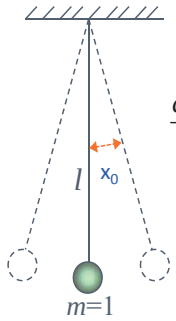
Research Center for Adaptive Data Analysis, National Central University, Taiwan
Institute of Physics, Academia Sinica, Taipei, Taiwan

E-mail: mcwu@ncu.edu.tw
Homepage: http://www.ncu.edu.tw/~mcwu/

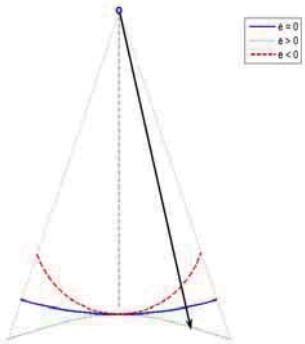
Plan

- **Introduction:**
  why time series?
- **Some interesting studies on:**
  financial data
  biomedical data
- **Conclusion**

Pendulum

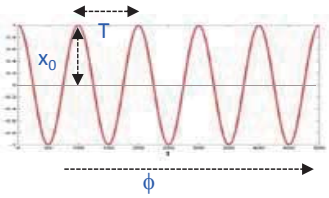


$$\frac{d^2x}{dt^2}+\frac{g}{l}x=0$$



Duffing equation

$$\frac{d^2x}{dt^2}+x+\varepsilon x^3=\gamma\cos\omega t$$



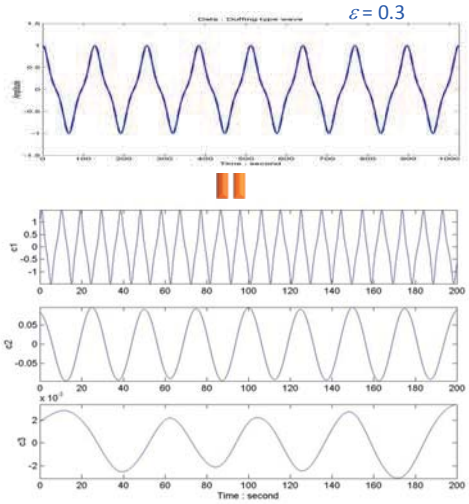
$$\frac{x}{x_0}=\cos\omega t,\quad \omega=\sqrt{\frac{g}{l}}$$

Amplitude: x_0

Period: $T=1/\omega$

Frequency: ω

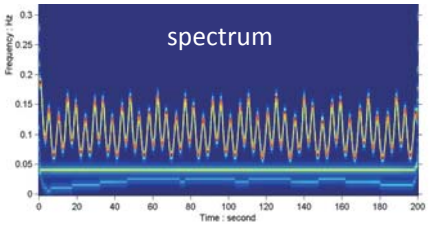
Phase: $\phi=\omega t$



Duffing type wave: $x = \cos(\omega t+\varepsilon \sin 2\omega t)$

For $\varepsilon << 1$, we can have

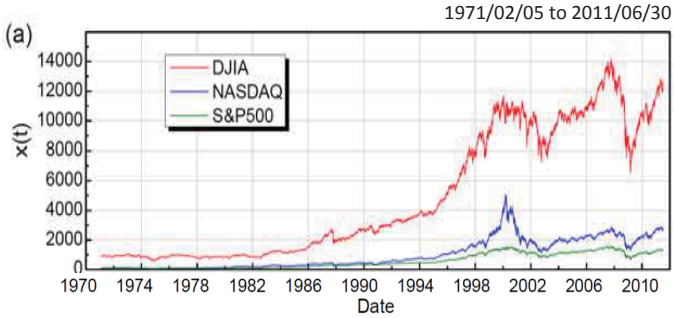
$$x(t)=\cos(\omega t+\varepsilon\sin 2\omega t)$$
$$=\cos\omega t\cos(\varepsilon\sin 2\omega t)-\sin\omega t\sin(\varepsilon\sin 2\omega t)$$
$$=\cos\omega t-\varepsilon\sin\omega t\sin 2\omega t+\ldots$$
$$=\left(1-\frac{\varepsilon}{2}\right)\cos\omega t+\frac{\varepsilon}{2}\cos 3\omega t+\ldots$$



Stock market indices

- **Dow Jones Industry Average (DJIA):** an average price of 30 companies, since 1928/10/01
- **NASDAQ:** consists of 1,197 stocks, since 1971/02/05
- **S&P500:** an average result of 500 companies, since 1950/01/03

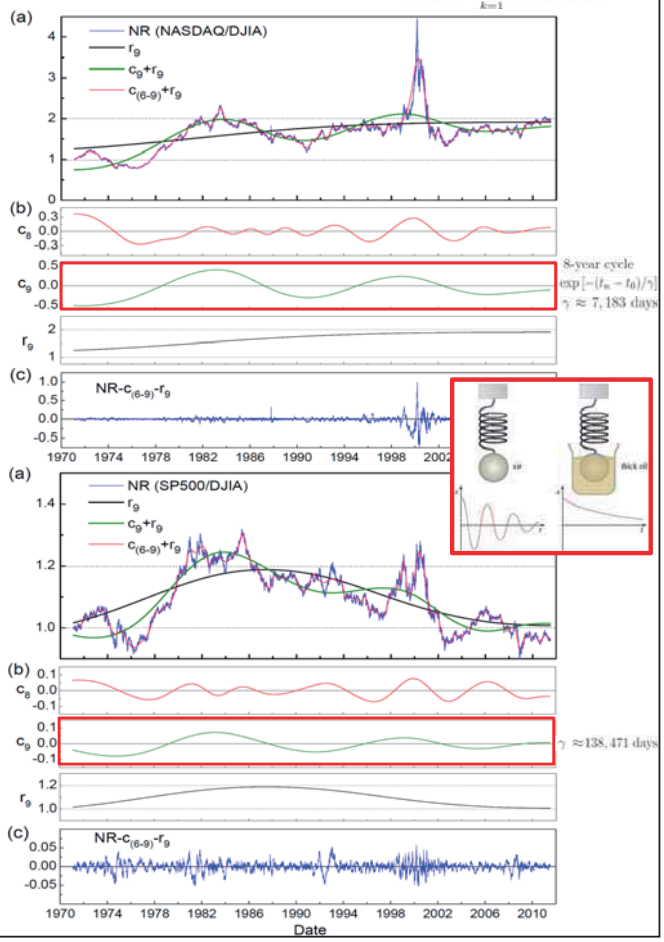
1971/02/05 to 2011/06/30



$$R_{ij}(t_n)=\frac{x_i(t_n)}{x_j(t_n)},\qquad NR_{ij}(t_n)\equiv\frac{R_{ij}(t_n)}{R_{ij}(t_0)}=\prod_{m=1}^{n}\frac{1+g_i(t_{m-1})}{1+g_j(t_{m-1})}.$$

gain of the index x_i $\quad g_i(t_n)=\dfrac{x_i(t_{n+1})-x_i(t_n)}{x_i(t_n)}$

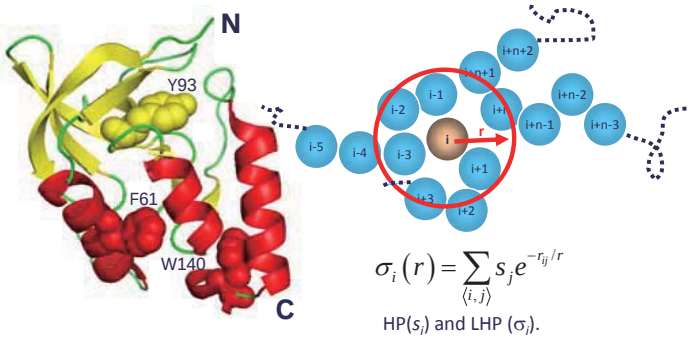
$$NR(t)=\sum_{k=1}^{n}c_k(t)+r_n(t).$$



M.-C. Wu, EPL 97, 48009 (2012).

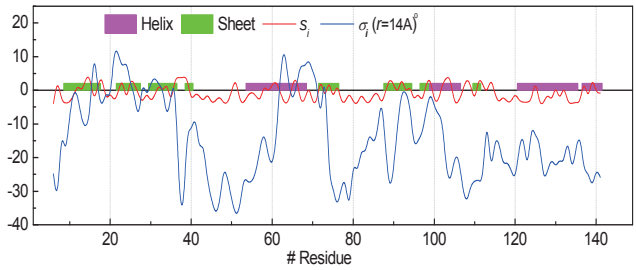
Hydrophobicity (HP) and local hydrophobicity (LHP)

HP(s_i) and LHP (σ_i) as functions of residues for Staphylococcal nuclease (Snase, 1EY0) estimated by the KD scale.

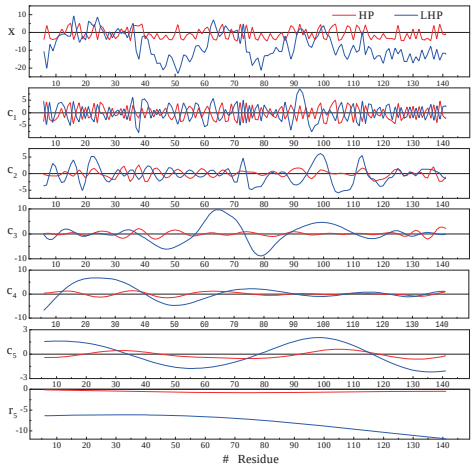


$$\sigma_i(r) = \sum_{\langle i,j \rangle} s_j e^{-r_{ij}/r}$$

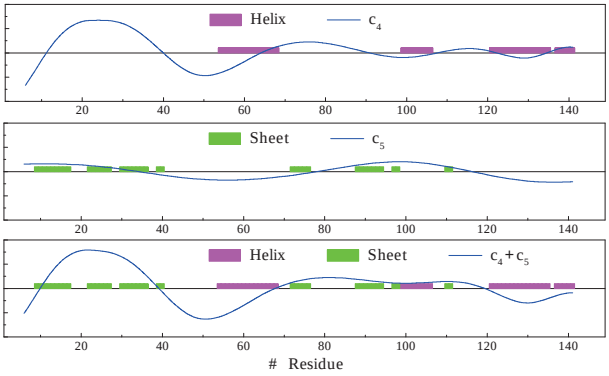
HP(s_i) and LHP (σ_i).



Decomposition of LHP



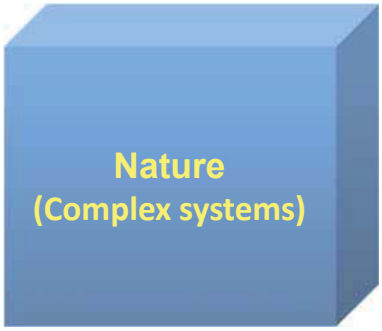
Secondary structures and IMFs c_4 and c_5 of LHP.



There are correlations between HP and LHP, and correlations between LHP and secondary structures. Folding information is imprinted in the sequence and is also preserved/carried by LHP.

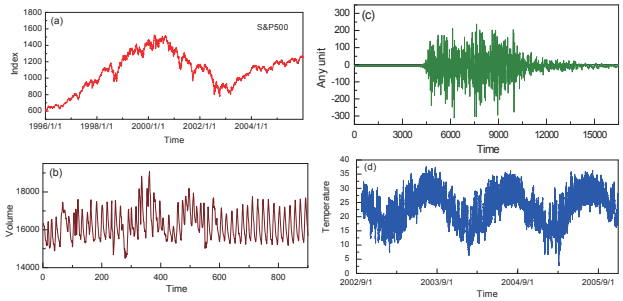
M.-C. Wu and T. Y. Tsong, J. Phys. Soc. 82, 114801 (2013).

Aspects of data analysis



Collect data

Analyze data to understand the complex systems



(a) S&P500 daily index, (b) respiratory time series, (c) seismic time series, and (d) temperature variation sampled by 5mins.

Tasks

- ☐ Physical processes are mostly nonstationary
- ☐ Physical Processes are mostly nonlinear
- ☐ Data from observations are invariably too short
- ☐ Physical processes are mostly non-repeatable

Development of useful data analysis approach is important

***Measures* of a time series**

- ☐ Amplitude
- ☐ Frequency
- ☐ Phase

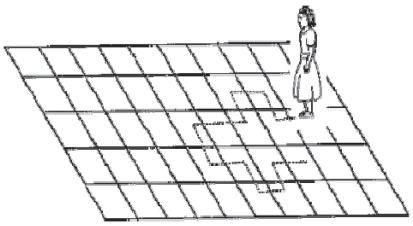
(also time series)

Time series analysis:

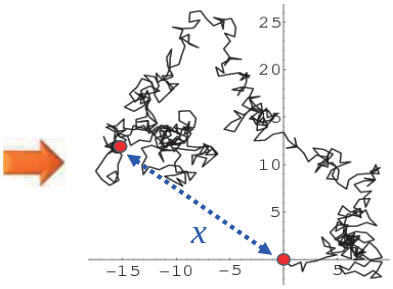
- ☐ Analysis based on **amplitude**: scaling analysis, variability, …
- ☐ Analysis based on **frequency**: Fourier spectrum, Hilbert spectrum, …
- ☐ Analysis based on their combinations: **complexity**, …
- ☐ Analysis based on instantaneous **phase**: phase distribution, phase correlation, …

Amplitude: scaling analysis

Random walk



http://scifun.chem.wisc.edu/WOP/RandomWalk.html



http://mathworld.wolfram.com/RandomWalk2-Dimensional.html

Indicator of walk efficiency

$$\langle x \rangle = l\sqrt{n} \propto n^{0.5}$$

<x> : end-to-end distance

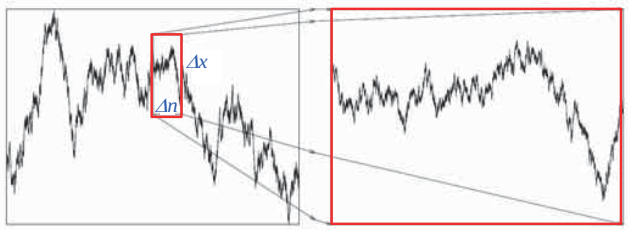
l : step length

n : number of steps

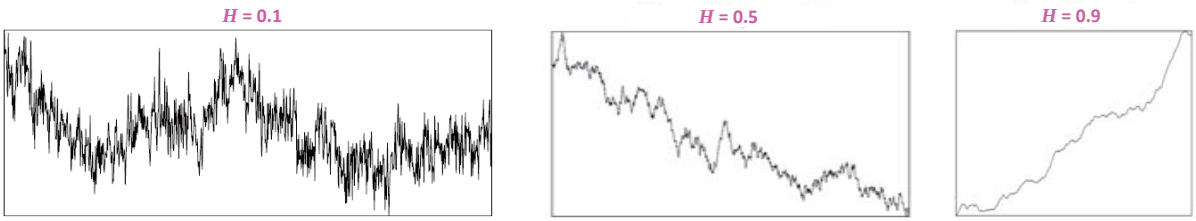
Fractional Brownian motion:

$$\Delta x \propto (\Delta n)^H$$ (H: Hurst exponent)

- Constant velocity: $0 \le H \le 1$
- Random walk: $H = 0.5$
- With acceleration: H can be larger than 1



Time series with different *parameters*:

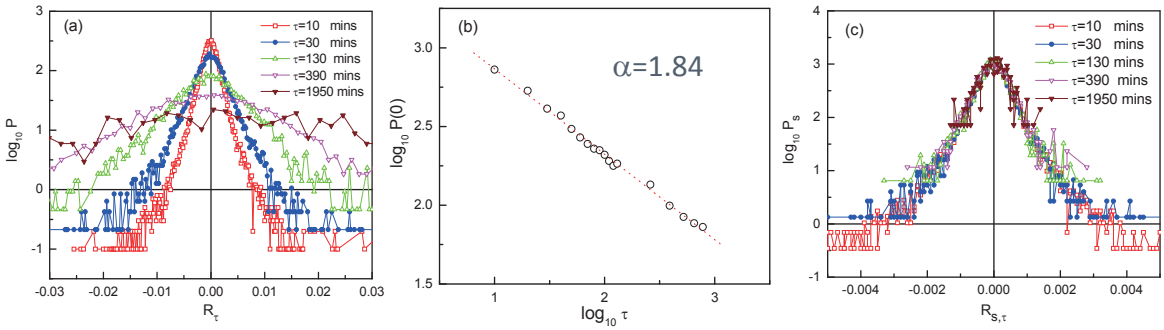


R. Blok, *Self-affine time series analysis* (2003)

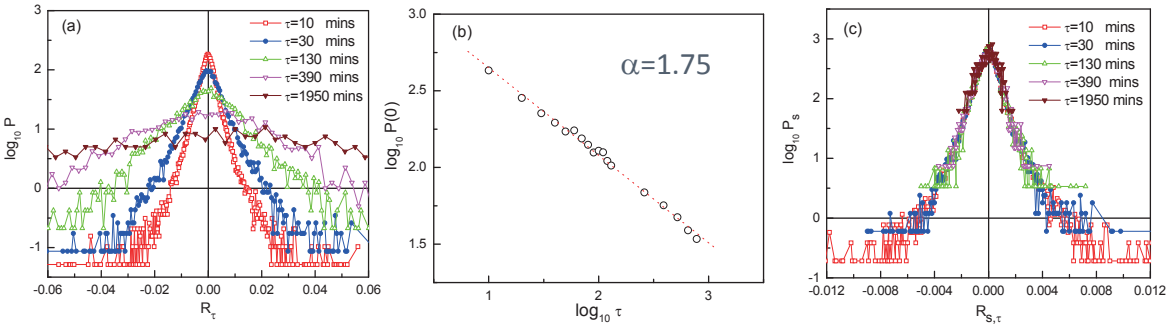
Scaling analysis

$$\log_{10} P(R_\tau = 0) = C - \frac{1}{\alpha}\log_{10}\tau, \qquad R_{s,\tau} = \frac{R_\tau}{\tau^{1/\alpha}}, \qquad P_s(R_{s,\tau}) = \frac{P(R_\tau)}{\tau^{-1/\alpha}}$$

DJIA intraday



NASDAQ intraday

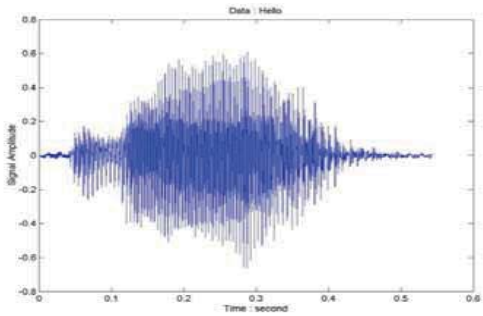


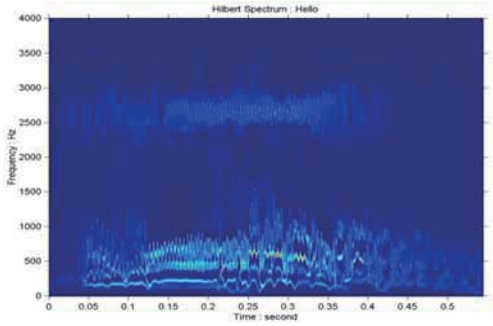
- α(=1.84, 1.75) satisfies the criterion of Levy's α stable distribution condition (α ≤ 2).
- Both DJIA and NASDAQ intraday returns have very nice scaling behaviors.

M.-C. Wu, *et al.*, Phys. Rev. E 73, 016118 (2006).

Frequency: spectral analysis

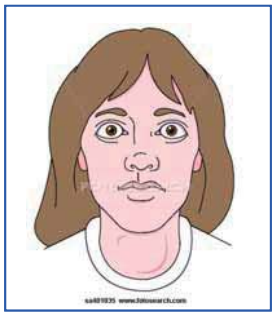
Spectrum: Speech analysis





Can be used for *voiceprint comparison.*

Power spectrum density (PSD)



Face of an adult female exhibiting the symptoms of hyperthyroidism.

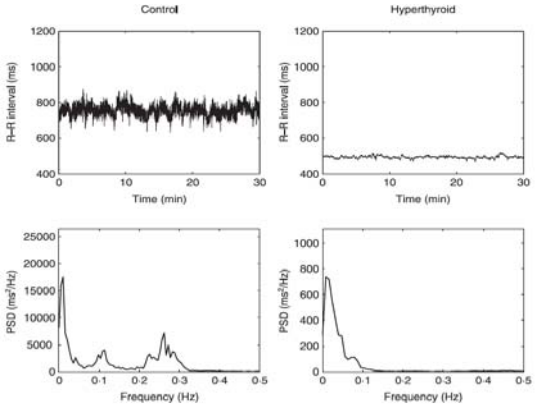


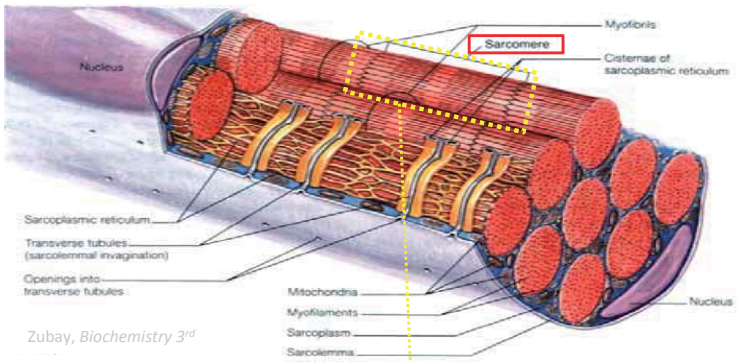
Fig. 1 R–R interval tachogram (upper) and the heart rate variability spectra (lower) in a normal control subject (left) and in a hyperthyroid patient (right). Note the different power spectral density (PSD) scale in the spectra due to the marked reduction in total power in the hyperthyroid patient.

The PSD of normal subjects is more *complex*.

J.-L. Chen, et al., Clinical Endocrinology 64, 611–616 (2006).

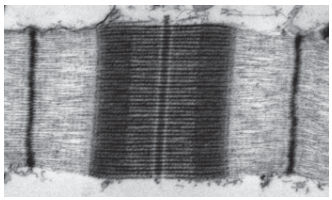
Example: *Amplitude + Frequency*

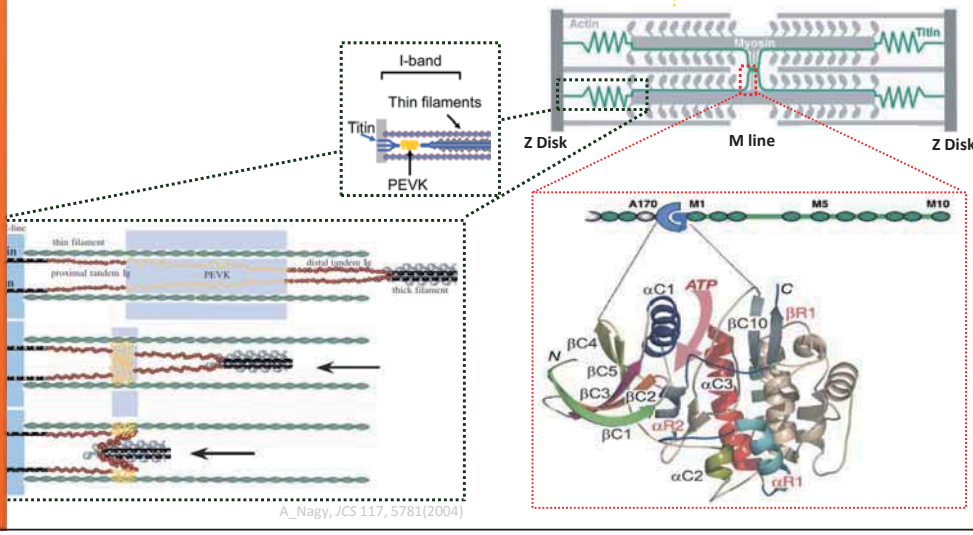
Striated muscle structure



Zubay, *Biochemistry 3rd*

Electron Microscope Image of Sarcomere



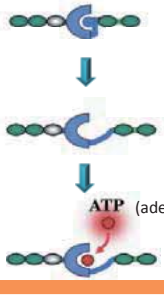


A_Nagy, *JCS* 117, 5781(2004)

Titin kinase --- *a digital force sensor*

MODEL:

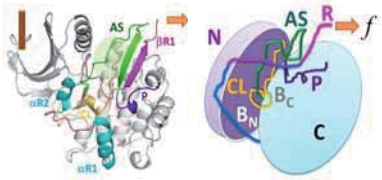
Stress-induced activation of titin kinase



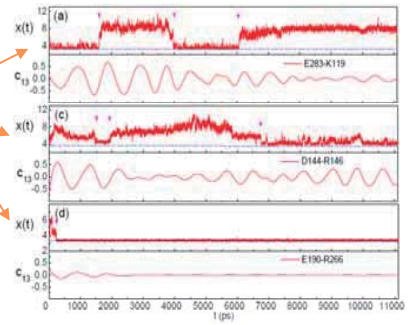
ATP (adenosine triphosphate)

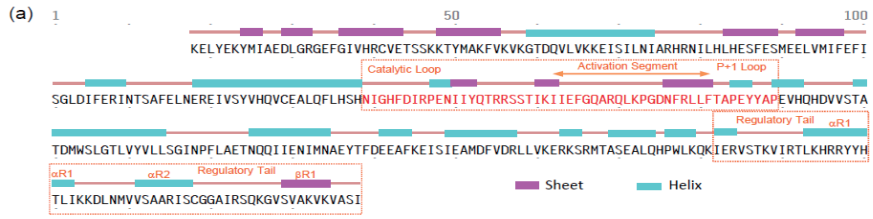
E. M. Puchner, et al. *PNAS* 105, 13385 (2008)

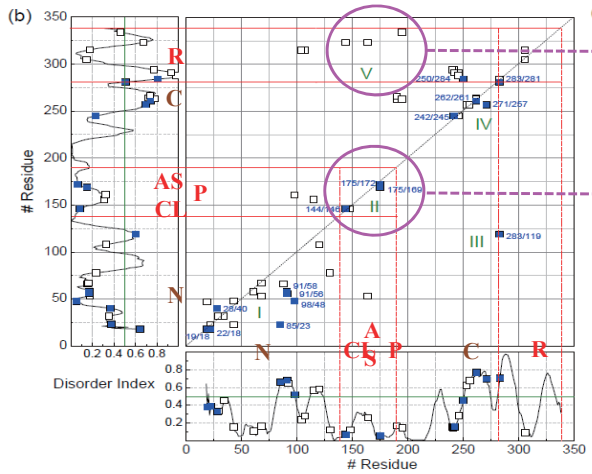
Mechanoactivation of titin kinase

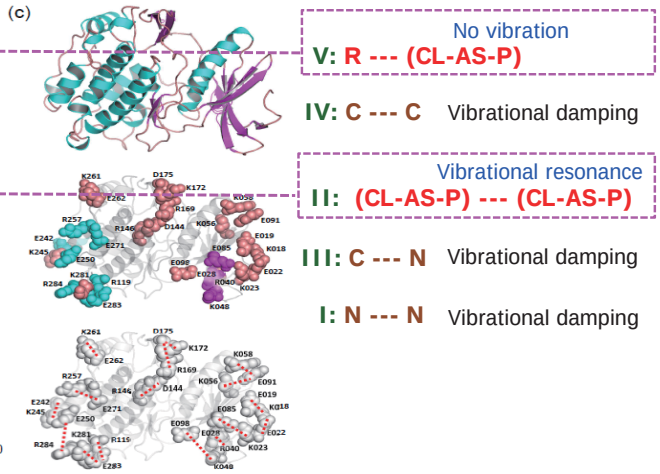


Groups	Ion pair connected regions	Mechanical properties
I, III, IV	N-N, C-C	Vibrational damping
II	AS-AS, CL-CL	Vibrational resonance
V*	R-R, R-AS, R-CL, AS-CL, CL-CL R-B_N, R-B_C, B_N-B_N, B_C-B_C	No vibration







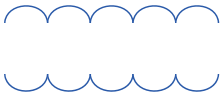


M.-C. Wu, J. G. Forbes, and K. Wang, Soft Matter 9, 9897-9906 (2013).

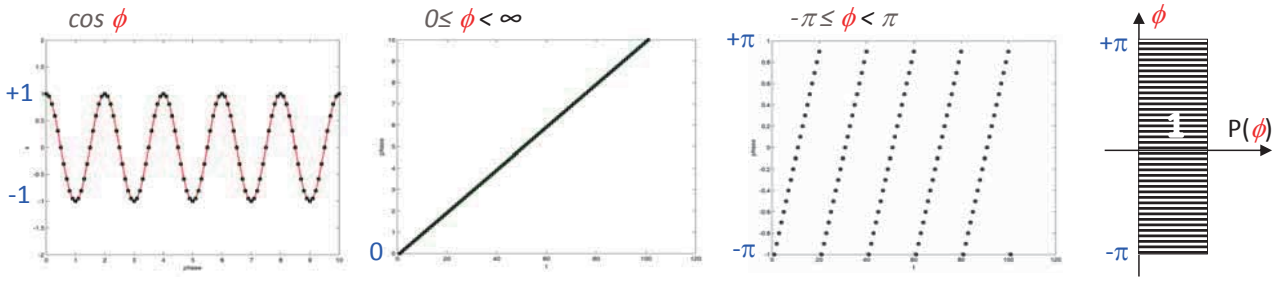
Phase: phase statistics

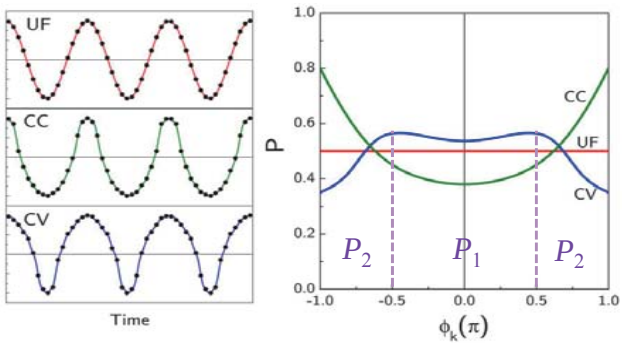
Analysis based on instantaneous phase: phase distribution, phase correlation, ...

Water wave ?



Idea: Phase catches some intrinsic properties of time series. It usually contains rich information about the structure of the data.



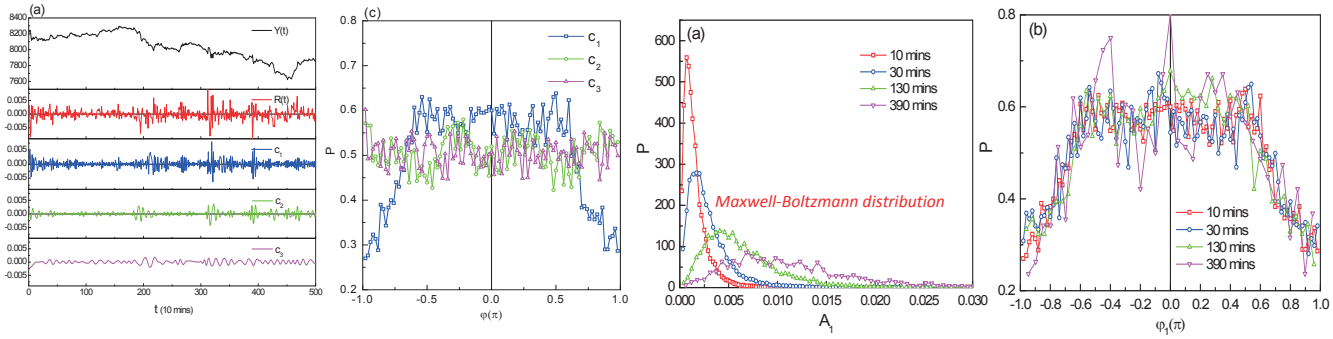


$$\chi = \langle P_1(\phi_1)\rangle - \langle P_2(\phi_1)\rangle$$

$P_1(\phi_1)$ is the PDF for the instantaneous phase ϕ_1 in the range $-0.5\pi \le \phi_1 \le 0.5\pi$, and $P_2(\phi_1)$ is for the phase in the range $\le -0.5\pi$ or $\ge 0.5\pi$.

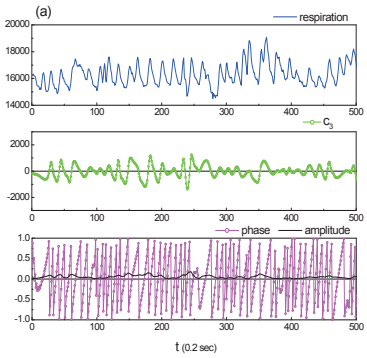
- CC: $\chi > \varepsilon$
- UF: $|\chi| \le \varepsilon$
- CV: $\chi < -\varepsilon$

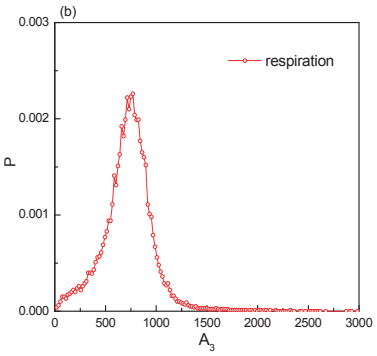
Distributions of amplitude and phase

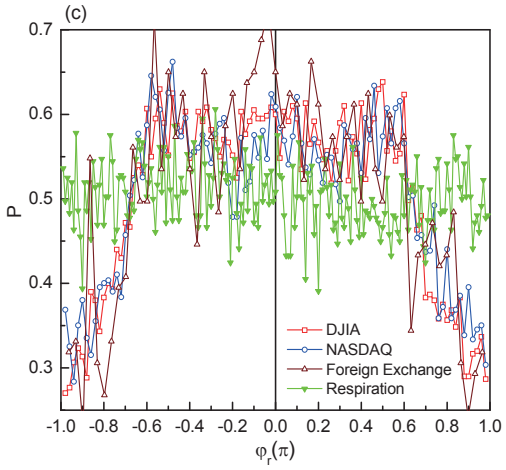


Most phases of the first-IMFs distribute at $-0.5\,\pi \le \phi \le 0.5\,\pi$.

This pattern existing in all time scale corresponds to abruptly change behaviors in time series of index, and indicates non-predictable and stochastic features of stock index.







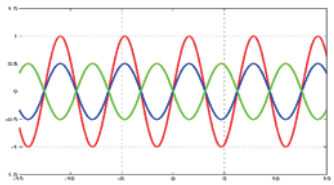
The phase distribution of a typical human respiratory time series is quite uniform.

M.-C. Wu, *et al.*, Phys. Rev. E 73, 016118 (2006).

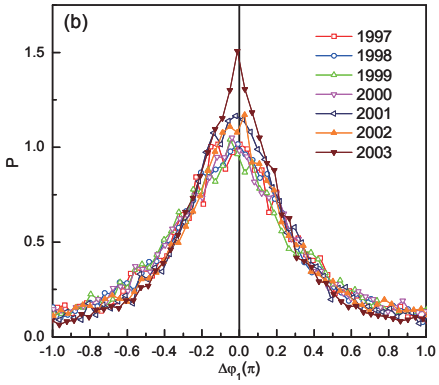
Phase differences between DJIA and NASDAQ

$$\Delta\phi_r = \phi_r(\text{NASDAQ}) - \phi_r(\text{DJIA})$$

Probability density distributions of phase differences between the first-IMFs of returns of DJIA and NASDAQ

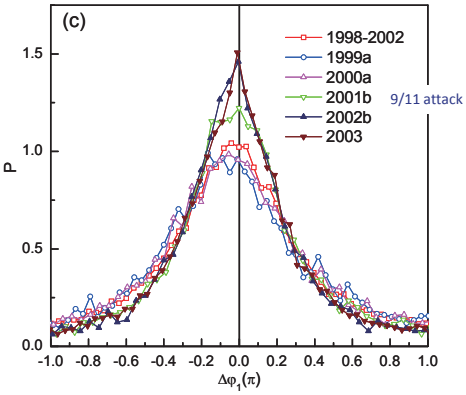


$$\Delta\phi(red - blue) = 0$$
$$\Delta\phi(red - green) = \pi$$

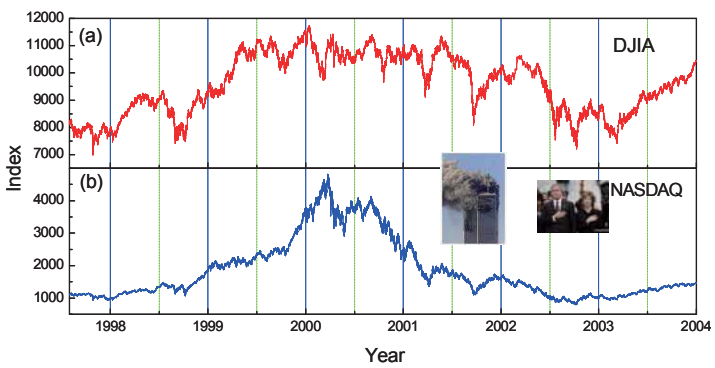


	1997	1998	1999	2000	2001	2002	2003
(b) skewness	-0.004	0.079	0.069	0.115	-0.025	0.003	0.026
kurtosis	0.907	0.843	0.728	0.779	1.326	0.934	1.990

More collective behaviors are found in year 2003.



	1998-2002	1999a	2000a	2001b	2002b	2003
(c) skewness	0.050	0.097	0.133	-0.050	-0.001	0.026
kurtosis	0.903	0.776	0.848	1.556	2.104	1.990



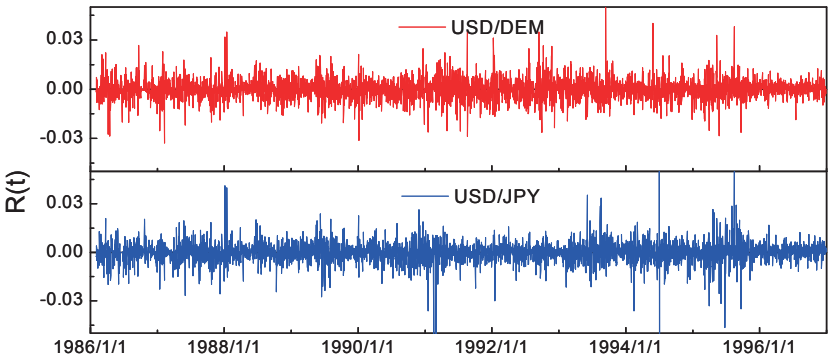
M.-C. Wu, *et al.*, Phys. Rev. E 73, 016118 (2006).

Foreign exchange data

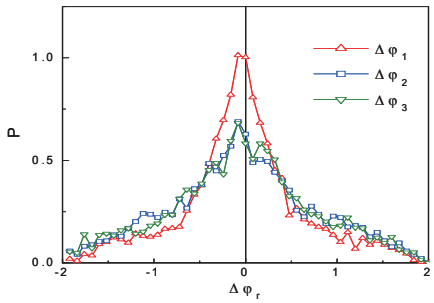
Empirical data of USD/DEM and USD/JPY exchange rates for a period from February 1 1986 to December 31 1996.

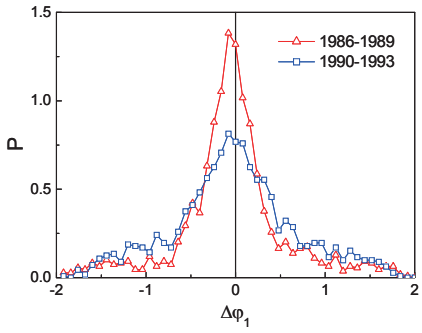
Intraday return: $R(t) = \frac{Y_c(t) - Y_o(t)}{Y_0(t)}$

Y(t): foreign exchange time series, subscript "o" denotes the opening price and "c" for the closing price.



$$\Delta\phi_r = \phi_r(\text{USD/JPY}) - \phi_r(\text{USD/DEM})$$





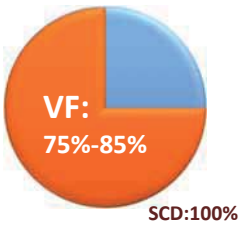
- Correlations are stronger in daily time scale than in longer time scales.
- The weaker correlation can be explained from the economic environments and government policies after the unification of East Germany and West Germany (1990.10.03) and in the era of bubble economy of Japan in early 1990s.

M.-C. Wu, Physica A 375, 633-642 (2007).

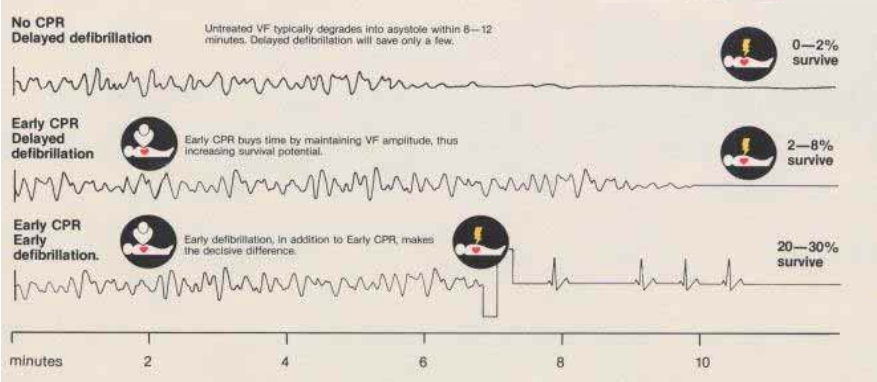
Example: Human Ventricular Fibrillation (VF)

More than 3 million people die annually from Sudden Cardiac Death (SCD) , with a survival rate of less than 1%.

Ventricular fibrillation (VF) is identified as a primary cause of SCD, accounting for 75% to 85% of sudden deaths. During VF, no blood is pumped, and blood circulation is terminated.



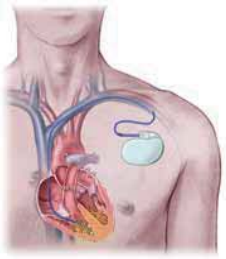
Zipes, D.P., et al. (2006) J. Am. Coll. Cardiol. 48, e247-346.
M.-C. Wu, E. Watanabe, Z. R. Struzik, C.-K. Hu, and Y. Yamamoto, Phys. Rev. E 80, 051917 (2009).



Timely defibrillation can often restore heartbeat in cases of VF.

http://www.defib-ecg-ekg-monitors.com/t_laerdal_2000_aed_defibrillators.asp

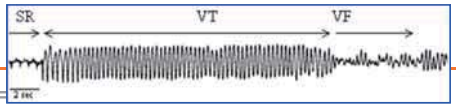
Automated External Defibrillator (AED) in Taiwan Taoyuan International Airport.



Implantable defibrillators has become the gold standard therapy for patients at a high risk of VF.

www.defibrillator-help.com/images/p2/1112.jpg

Empirical Data



Code	Age	Sex	Underlying disease[a]	Ventricular arrhythmia[b]	Outcome	VTA analysis time	Δt (sec)[c]	χ [d]	Pattern[e]
S_{01}	46	M	IHD	VT→VF	Dead	7:37:40-	9.324	0.044	CV
S_{02}	68	M	IHD	VF	Dead	22:59:51–23:01:31	100.992	0.096	CV
S_{03}	67	M	IHD	VT→VF	Dead	19:42:10-	405.000	−0.100	CC
S_{04}	97	F	CAVB	VT→VF		3:56:31–3:57:45	73.992	0.030	CV
				VT→VF		14:04:49–14:05:07	19.992	−0.094	CC
				VT→VF	Dead	21:38:56–21:40:19	82.992	−0.010	UF
S_{05}	80	M	IHD	VT→VF		10:32:03-	228.000	0.099	CV
				VT→VF		10:37:27–10:37:50	25.000	0.052	CV
				VT→VF		10:38:29–10:39:17	49.000	−0.093	CC
				VT→VF		10:39:54–10:40:33	40.000	0.075	CV
				VT→VF	Dead	11:05:58–11:09:30	213.000	0.068	CV
S_{06}	22	F	LQT	TdP		1:11:50–1:12:10	19.992	−0.094	CC
				VT→VF	Alive	3:44:07–3:46:01	113.992	0.006	UF
S_{07}	75	F	CAVB	VT→VF	Alive	14:53:00-	64.208	−0.006	UF
S_{08}	80	M	IHD	VT→VF	Dead	12:22:58–12:26:10	191.992	−0.041	CC
S_{09}	70	M	HTN	VT→VF	Alive	15:53:52–15:55:25	92.992	−0.039	CC
S_{10}	67	M	IHD	VT→VF	Dead	23:07:50–23:26:00	1089.992	0.052	CV
S_{11}	71	M	IHD	VT→VF		16:49:30–16:50:40	69.992	−0.045	CC
				VT→VF		00:24:05–0:25:40	94.992	−0.044	CC
				VT→VF		3:22:15–3:25:30	194.992	−0.082	CC
				VT→VF	Alive	6:05:10–6:07:50	159.992	−0.047	CC
S_{12}	66	M	IHD	VT→VF	Dead	12:38:31–12:52:40	848.992	0.023	UF
S_{13}	56	M	CAVB	TdP	Alive	20:01:38–20:02:08	29.992	−0.083	CC
S_{14}	88	M	IHD	VT	Dead	4:44:30-	114.992	0.005	UF
S_{15}	51	F	HTN	TdP	Alive	12:36:24–12:37:15	50.992	−0.021	UF
S_{16}	75	M	IHD	VT		14:47:40–15:13:45	1564.992	−0.044	CC
				VT→VF	Dead	15:50:10–16:09:00	1129.992	−0.016	UF
S_{17}	43	M	ND	VT→VF	Dead	18:41:06–18:50:38	492.800	−0.037	CC

[a]Ischemic heart disease (IHD); complete atrioventricular block (CAVB); long-QT (LQT) syndrome; hypertension (HTN); not detected (ND).
[b]VT; (VF); torsades de pointes (TdP).
[c]Δt is the available data time interval for analysis.
[d]χ is defined in Eq. (8).
[e]CC; CV; UF.

Mobile Nurse

Data collected by Fujita Healthy University Hospital in Toyoake, Japan.

17 VF patients:
6 survivors,
11 non-survivors.

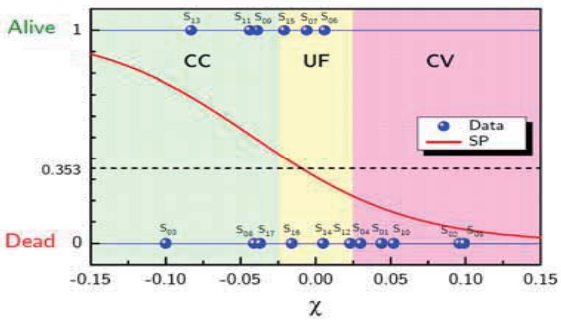
Survival rate: 35.3%

M.-C. Wu, E. Watanabe, Z. R. Struzik, C.-K. Hu, and Y. Yamamoto, Phys. Rev. E 80, 051917 (2009).

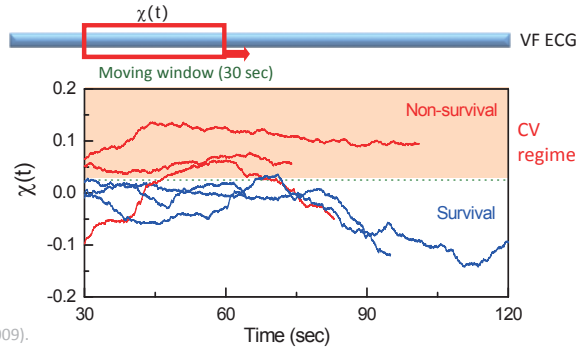
Phase distribution patterns of human VF

Pattern	No. of Cases	Survival	Non-Survival
CC	5	3 (60%)	2 (40%)
UF	7	3 (43%)	4 (57%)
CV	5	0 (0%)	5 (100%)

CV type VF is likely the non-self-terminating VF !

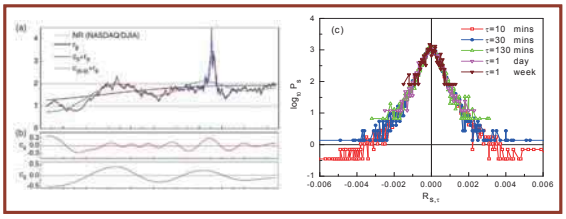


Analysis on the VF intervals lasting more than 30 sec

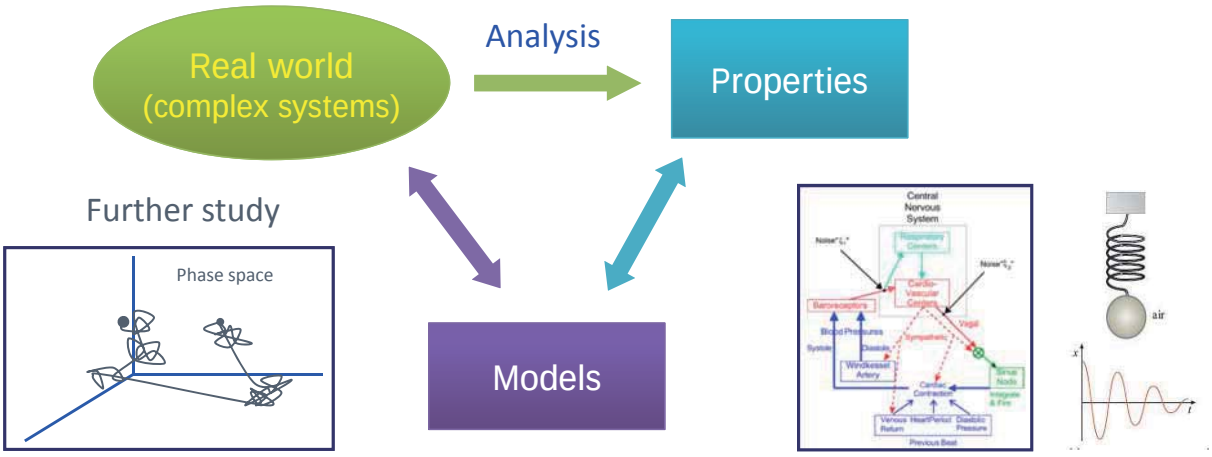


M.-C. Wu, E. Watanabe, Z. R. Struzik, C.-K. Hu, and Y. Yamamoto, Phys. Rev. E 80, 051917 (2009)
M.-C. Wu, Z. R. Struzik, E. Watanabe, Y. Yamamoto, and C.-K. Hu, AIP Conf. Proc. 922, 573-576 (2007).
M.-C. Wu and N. E. Huang, Advance Biosignal Processing, Amine NAIT-ALI Ed., Chapter 19, Springer (2009).

Conclusion: Way of Science



Understand the systems by analyzing the corresponding data



Time series analysis usually is the first step to understand how Nature works.

References

- M.-C. Wu, M.-C. Huang, H.-C. Yu, and T.C. Chiang, *Phase distribution and phase correlation of financial time series*, Phys. Rev. E 73, 016118 (2006).
- M.-C. Wu and C.-K. Hu, *Empirical mode decomposition and synchrogram approach to cardiorespiratory synchronization*, Phys. Rev. E 73, 051917 (2006).
- M.-C. Wu, *Phase correlation of foreign exchange time series*, Physica A 375, 633-642 (2007).
- M.-C. Wu, *Phase statistics approach to time series analysis*, J. Korean Phys. Soc. 50, 304-312 (2007).
- M.-C. Wu, *Phase statistics approach to physiological and financial time series*, AAPPS Bulletin 17, 21-26 (2007).
- M.-C. Wu, Z. R. Struzik, E. Watanabe, Y. Yamamoto, and C.-K. Hu, *Temporal evolution for the phase histogram of ECG during human ventricular fibrillation*, AIP Conf. Proc. 922, 573-576 (2007).
- M.-C. Wu, *Phase statistics approach to financial time series*, Phys. Bimonth. 30, 257-264 (2008).
- M.-C. Wu and C.-K. Hu, *Application of Empirical Mode Decomposition to Cardiorespiratory Synchronization*, In Complex Dynamics in Physiological Systems, S. K. Dana et al. Ed., Chapter 11, p.167-181, Springer (2009).
- M.-C. Wu and N. E. Huang, *Biomedical data processing using HHT: A review*, In Advance Biosignal Processing, Amine NAIT-ALI Ed., Chapter 19, Springer (2009).
- M.-C. Wu, E. Watanabe, Z. R. Struzik, C.-K. Hu, and Y. Yamamoto, *Phase statistics approach to human ventricular fibrillation*, Phys. Rev. E 80, 051917 (2009).
- M.-C. Wu, *Damped oscillations in the ratios of stock market indices*, EPL 97, 48009 (2012).
- M.-C. Wu, J. G. Forbes, and K. Wang, *Correlated vibrations in ion-pair dynamics in mechanoactivation identifies functional domains of force-dependent titin kinase*, Soft Matter 9, 9897-9906 (2013).
- M.-C. Wu and T. Y. Tsong, *Local hydrophobicity in protein secondary structure formation*, J. Phys. Soc. Japan 82, 114801 (2013).

Available at: http://www.ncu.edu.tw/~mcwu/