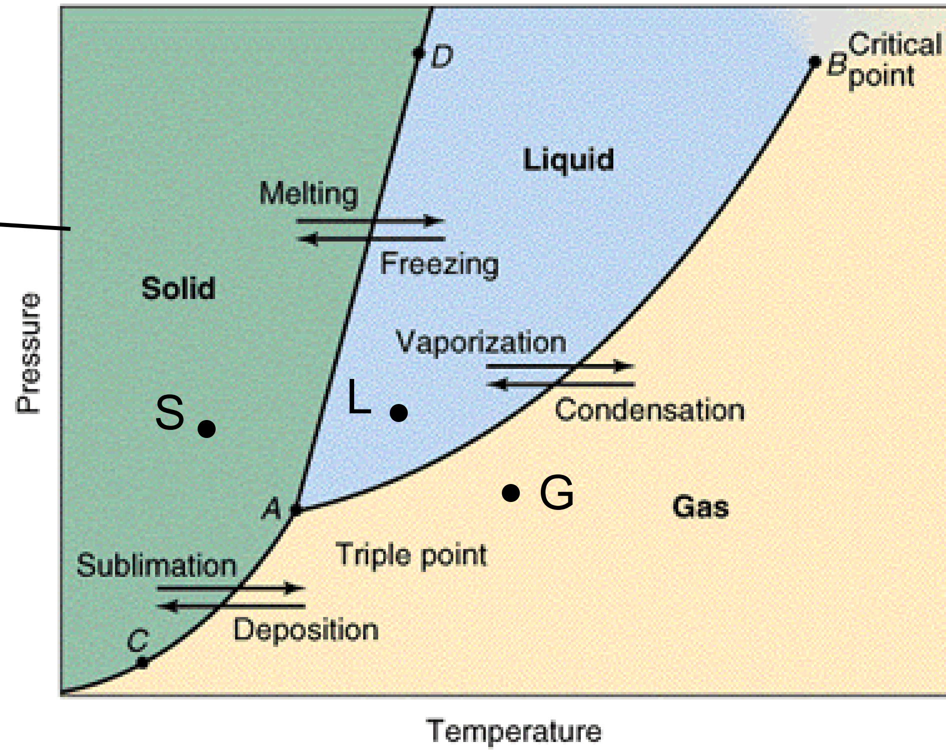
NANYANG TECHNOLOGICAL UNIVERSITY
新加坡南洋理工大学

# Time Series Approaches to Understanding Physical Phenomena

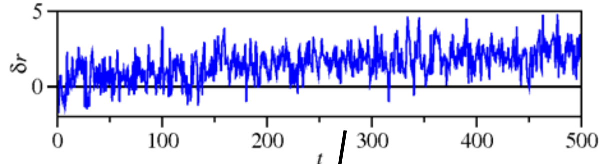
CHEONG Siew Ann
**张 寿 安**
**cheongsa@ntu.edu.sg**
**http://www1.spms.ntu.edu.sg/~cheongsa/**

# [Statistical] Physics in a Drop of Water

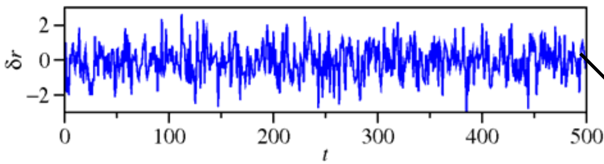



- Macroscopic order parameters differentiate
  - Solid (S)
  - Liquid (L)
  - Gas (G)

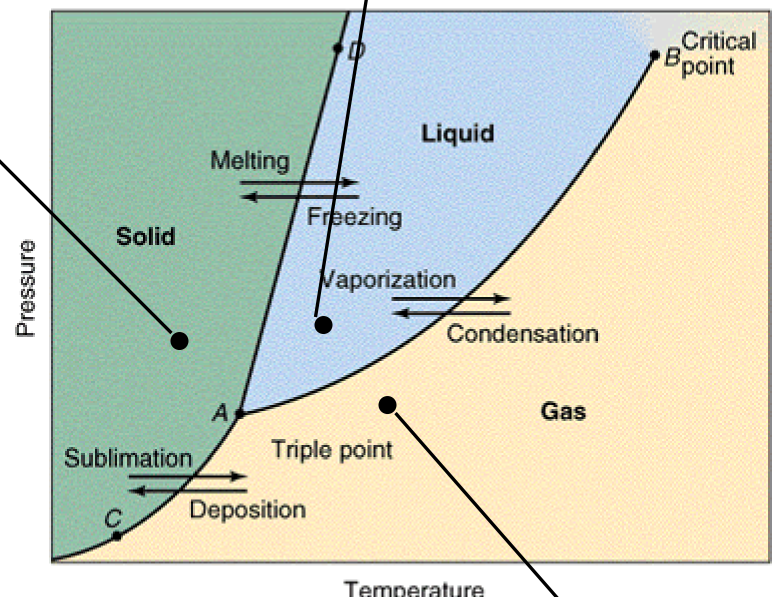
# [Statistical] Physics in a Drop of Water



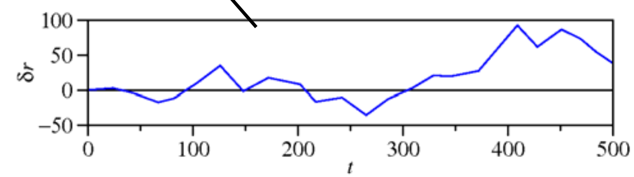

diffusive trajectories,
$\delta r^2$ increases with time




$\delta r$ fluctuates about 0,
$\delta r^2 = \alpha T$ time-independent

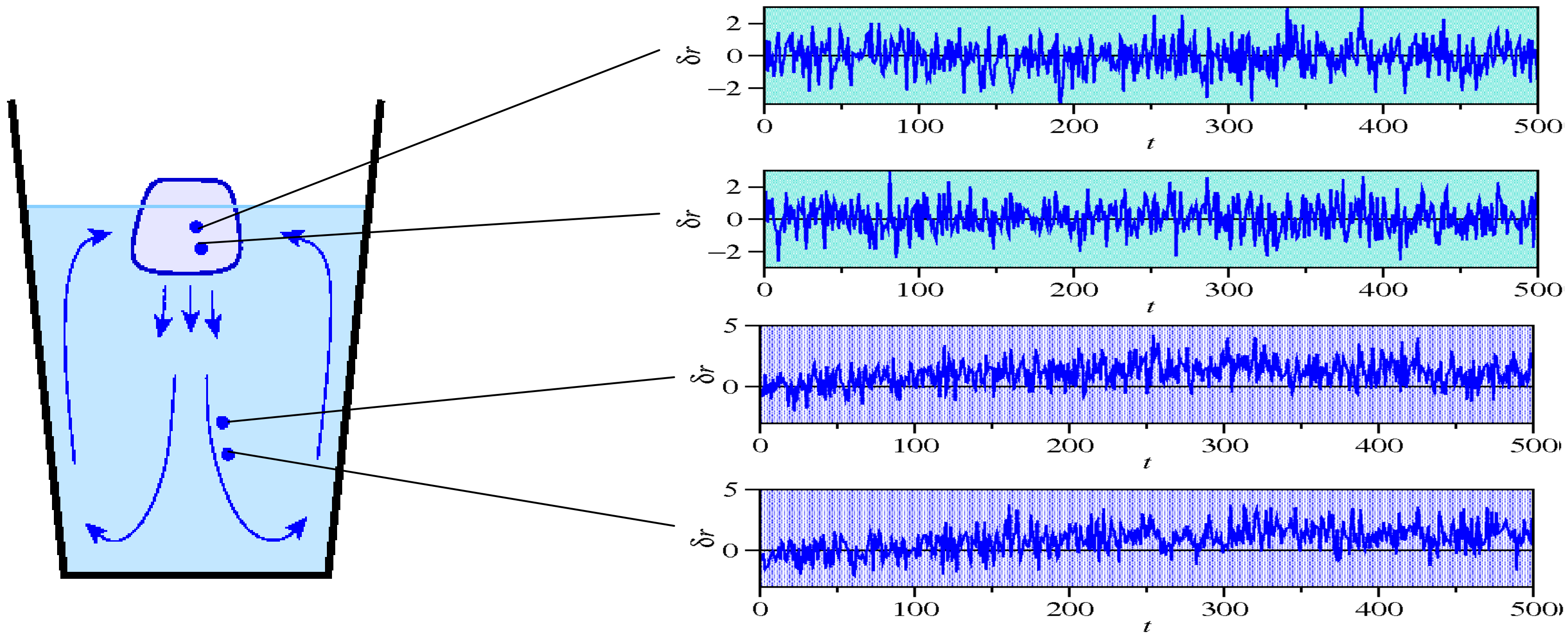



ballistic trajectories,
infrequent collisions




- S, L, G time series distinguishable
- S, L, G phase within single time series distinguishable

# [Statistical] Physics in a Drop of Water

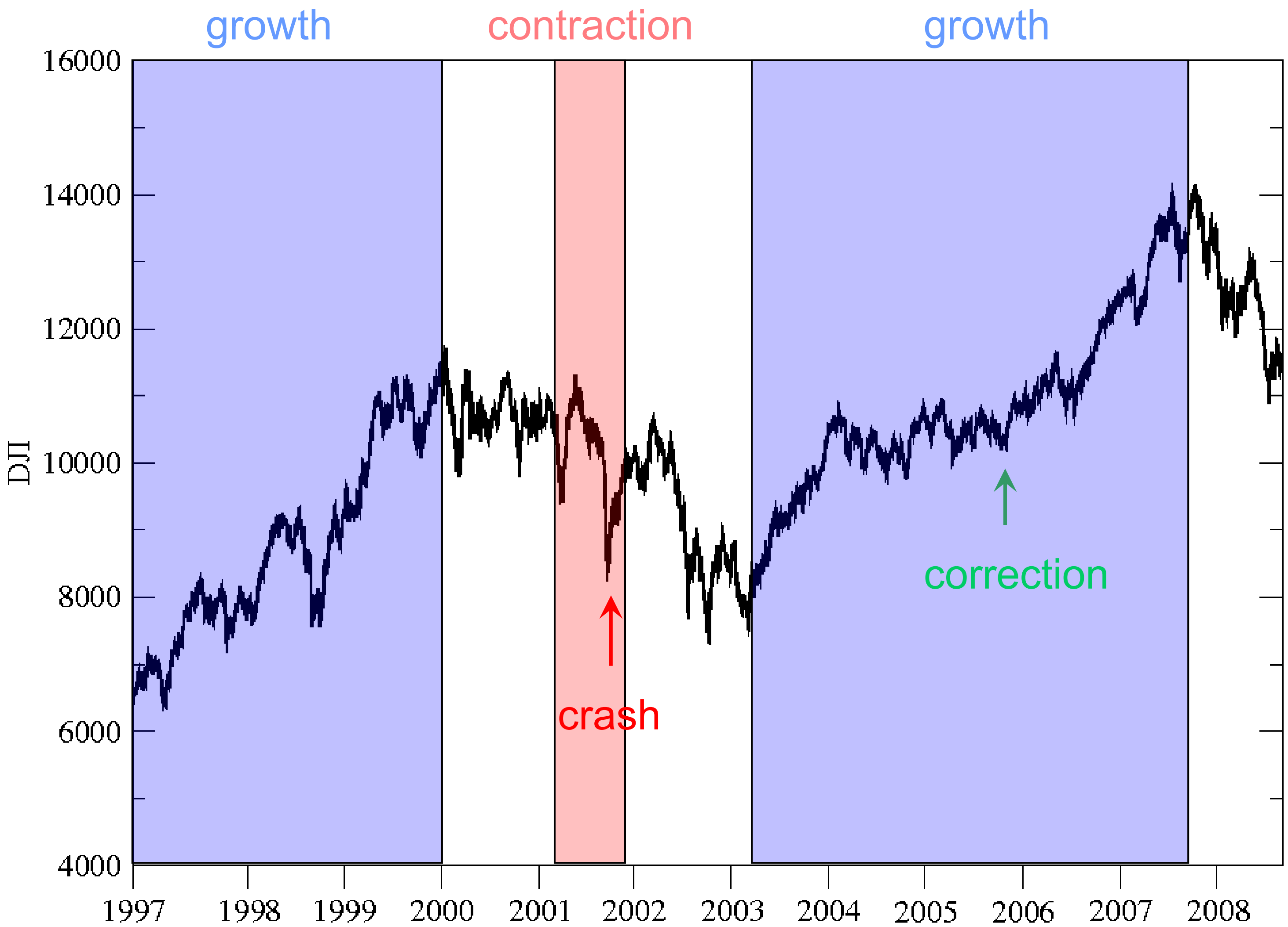



- Group statistically similar time series
- Discover presence of different phases

# Time Series Approaches

- Time Series Segmentation
  - Discover number/type of macroscopic phases
  - Discover lifetimes of macroscopic phases
  - Discover time scales of transitions between macroscopic phases
- Time Series Clustering
  - Discover effective mesoscopic variables in given time window
  - Discover slow time evolution of effective variables by sliding time window

# Time Series Segmentation



# Time Series Segmentation

- Assume non-stationary time series **x** = ($x_1$, $x_2$, …, $x_N$) consists of $M$ stationary segments
- In segment $m$, data points drawn from ($\mu_m$, $\sigma_m^2$) Gaussian distribution
- Recursive segmentation
  - One time series → two segments
  - Each segment → two subsegments
  - Iterate + optimize
  - Terminate

# Time Series Segmentation

- Single-segment likelihood for **x** = ($x_1$, $x_2$, …, $x_N$)

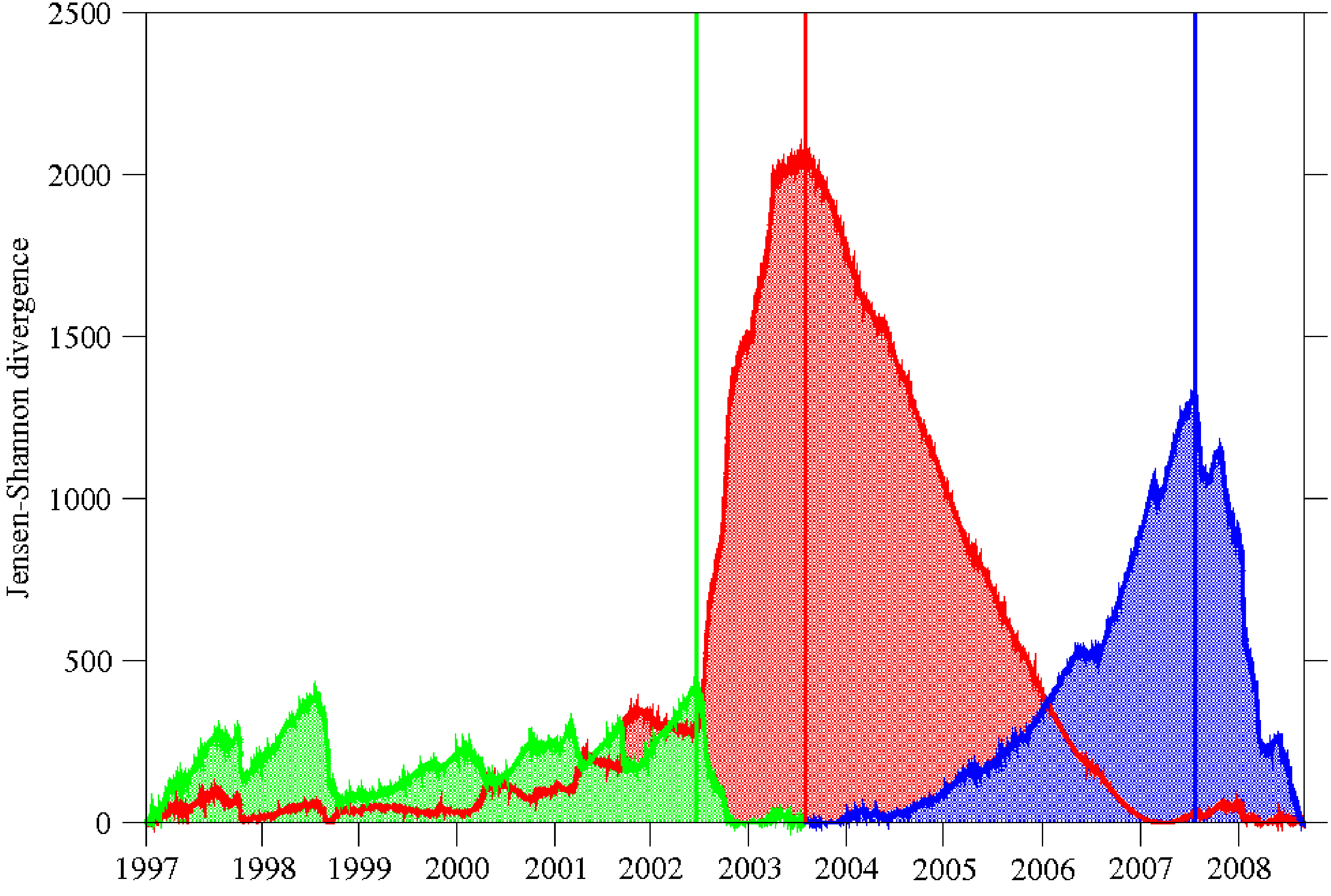
$$L_1 = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[ -\frac{(x_i - \mu)^2}{2\sigma^2} \right]$$

- Two-segment likelihood for **x** = ($x_1$, …, $x_t$, $x_{t+1}$, …, $x_N$)

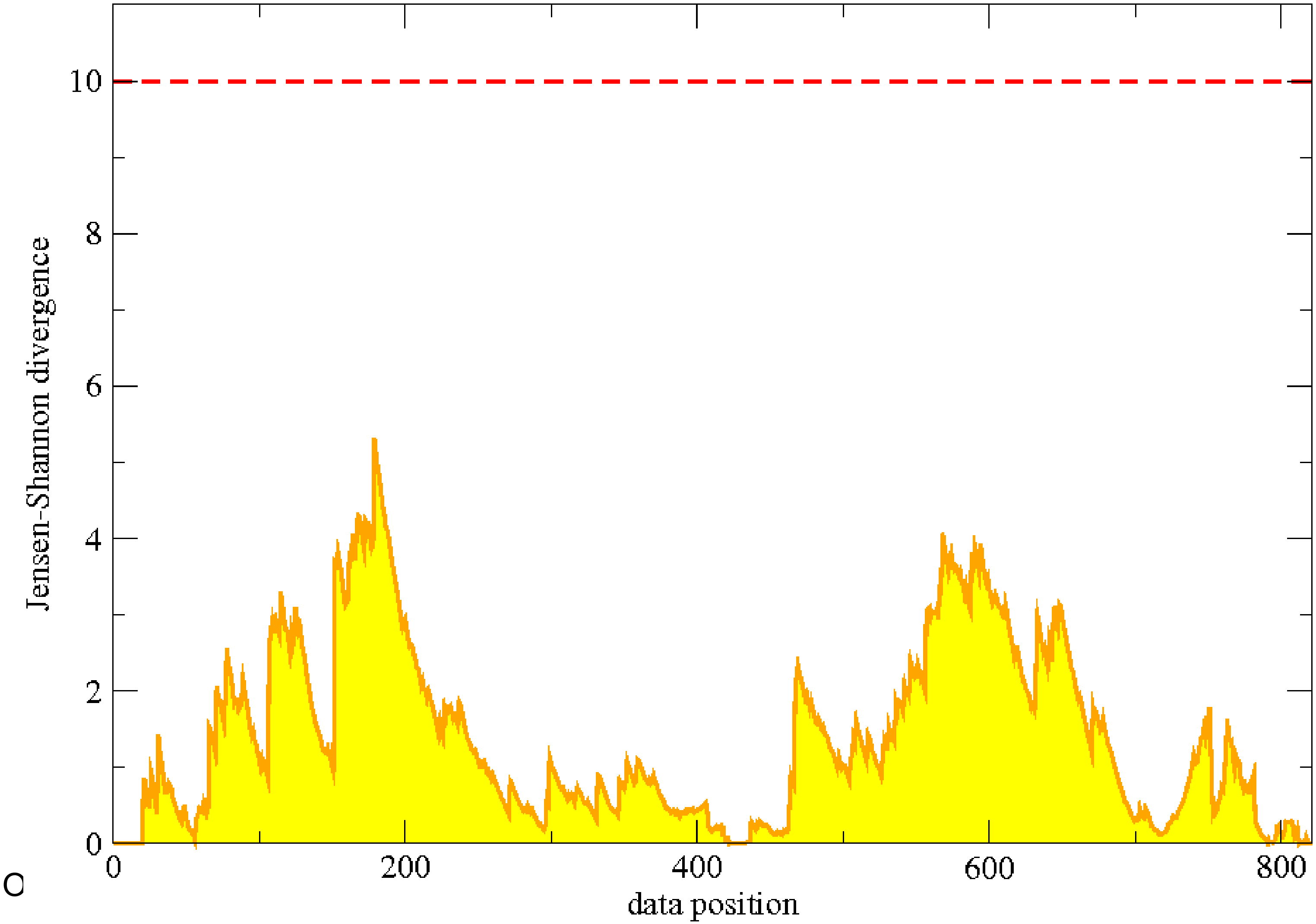
$$L_2(t) = \prod_{i=1}^{t} \frac{1}{\sqrt{2\pi\sigma_L^2}} \exp\left[ -\frac{(x_i - \mu_L)^2}{2\sigma_L^2} \right] \prod_{i=t+1}^{N} \frac{1}{\sqrt{2\pi\sigma_R^2}} \exp\left[ -\frac{(x_i - \mu_R)^2}{2\sigma_R^2} \right]$$

- ML estimates $\hat{\mu}, \hat{\mu}_L, \hat{\mu}_R, \hat{\sigma}^2, \hat{\sigma}_L^2, \hat{\sigma}_R^2$
- Jensen-Shannon divergence $\Delta(t) = \ln \frac{L_2(t)}{L_1} \geq 0$

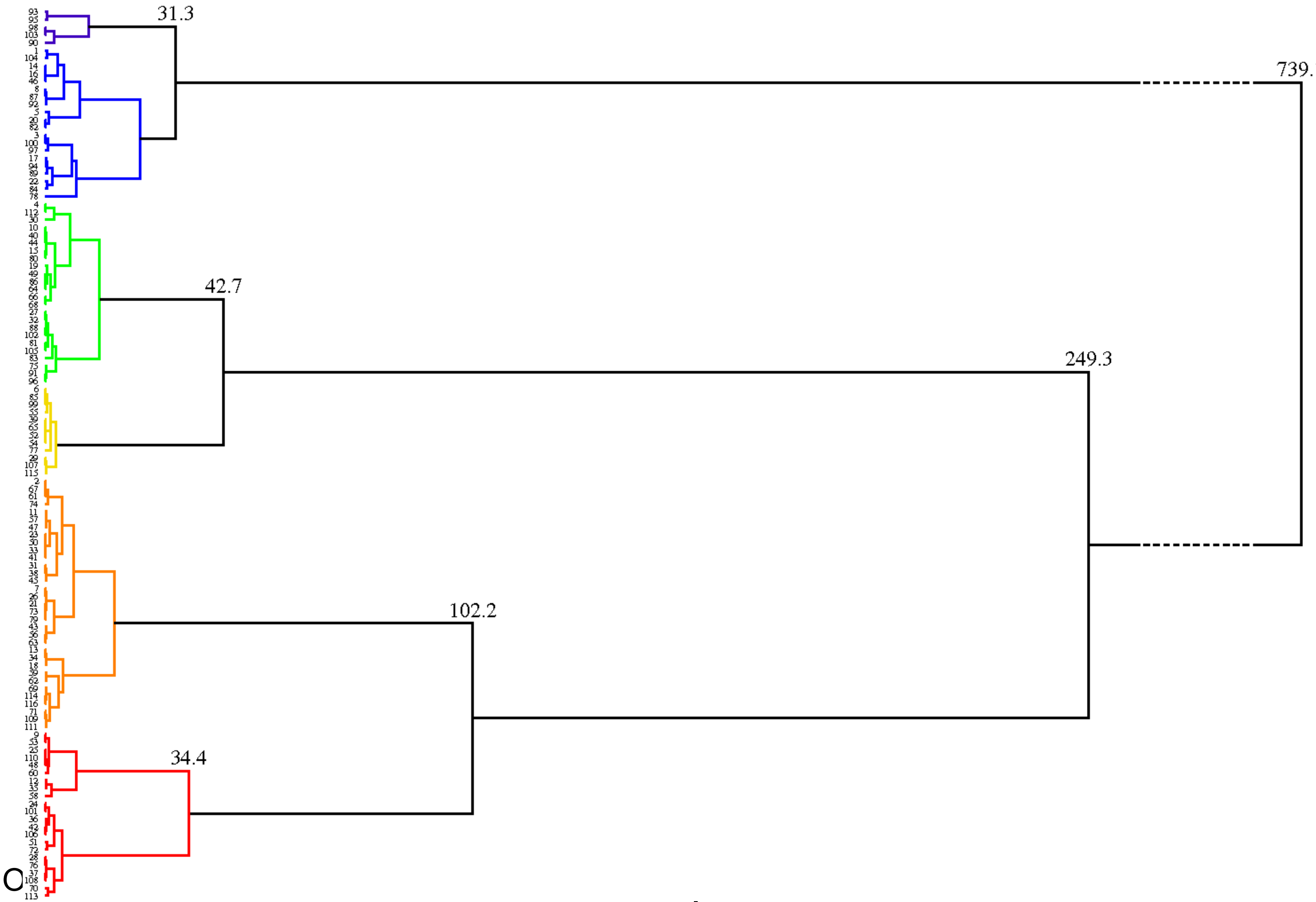
# Time Series Segmentation



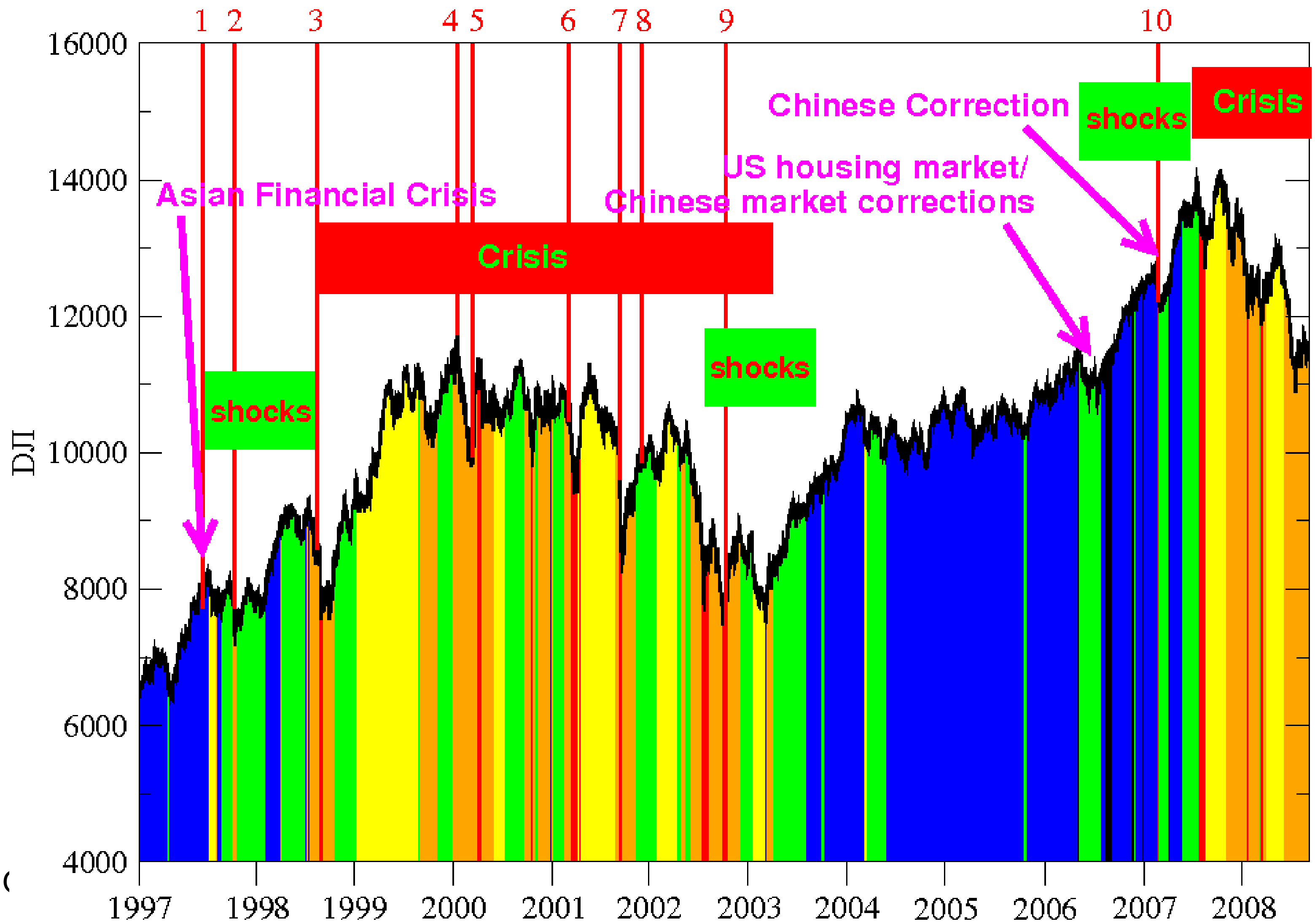
# Time Series Segmentation



# Time Series Segmentation



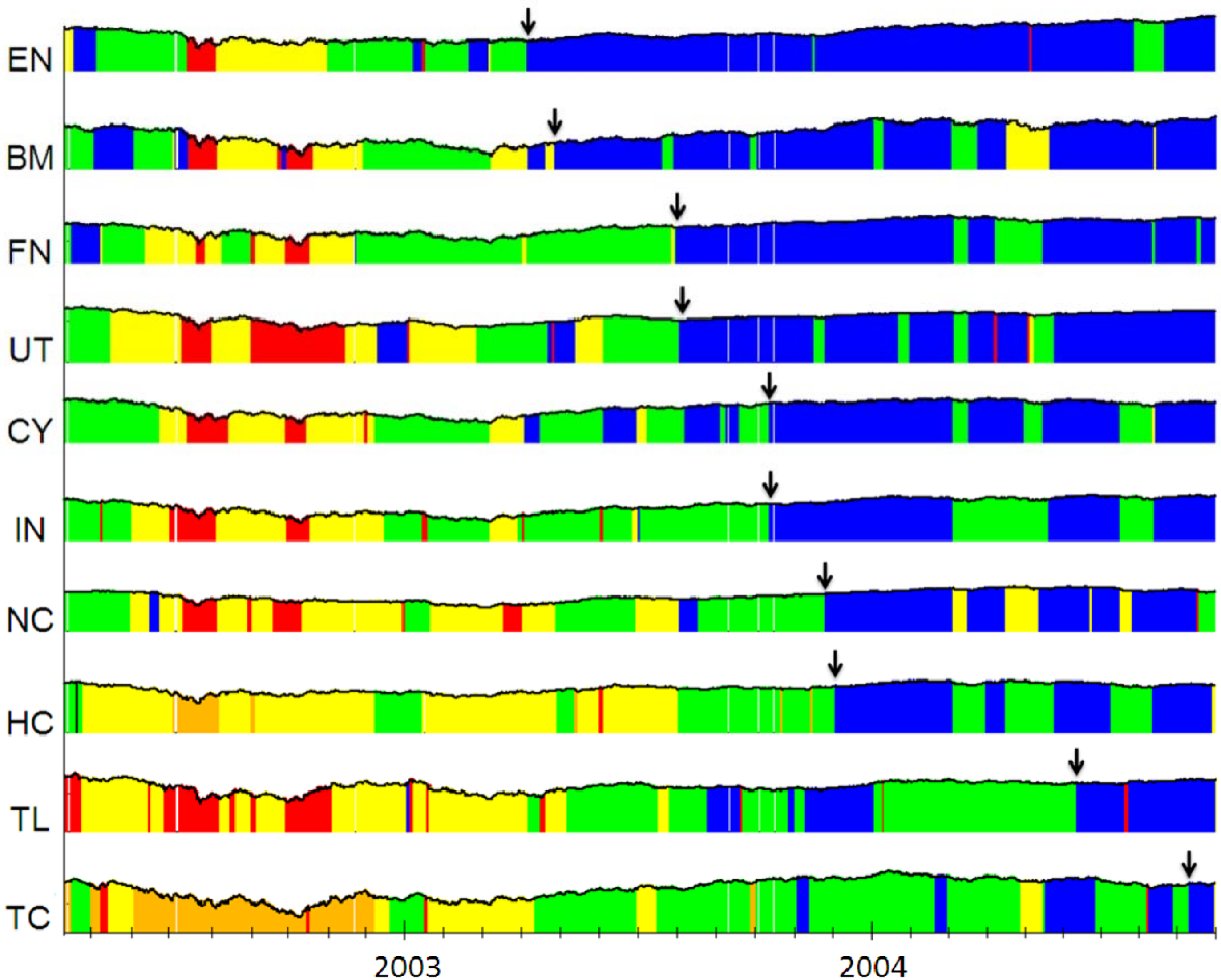
# Time Series Segmentation



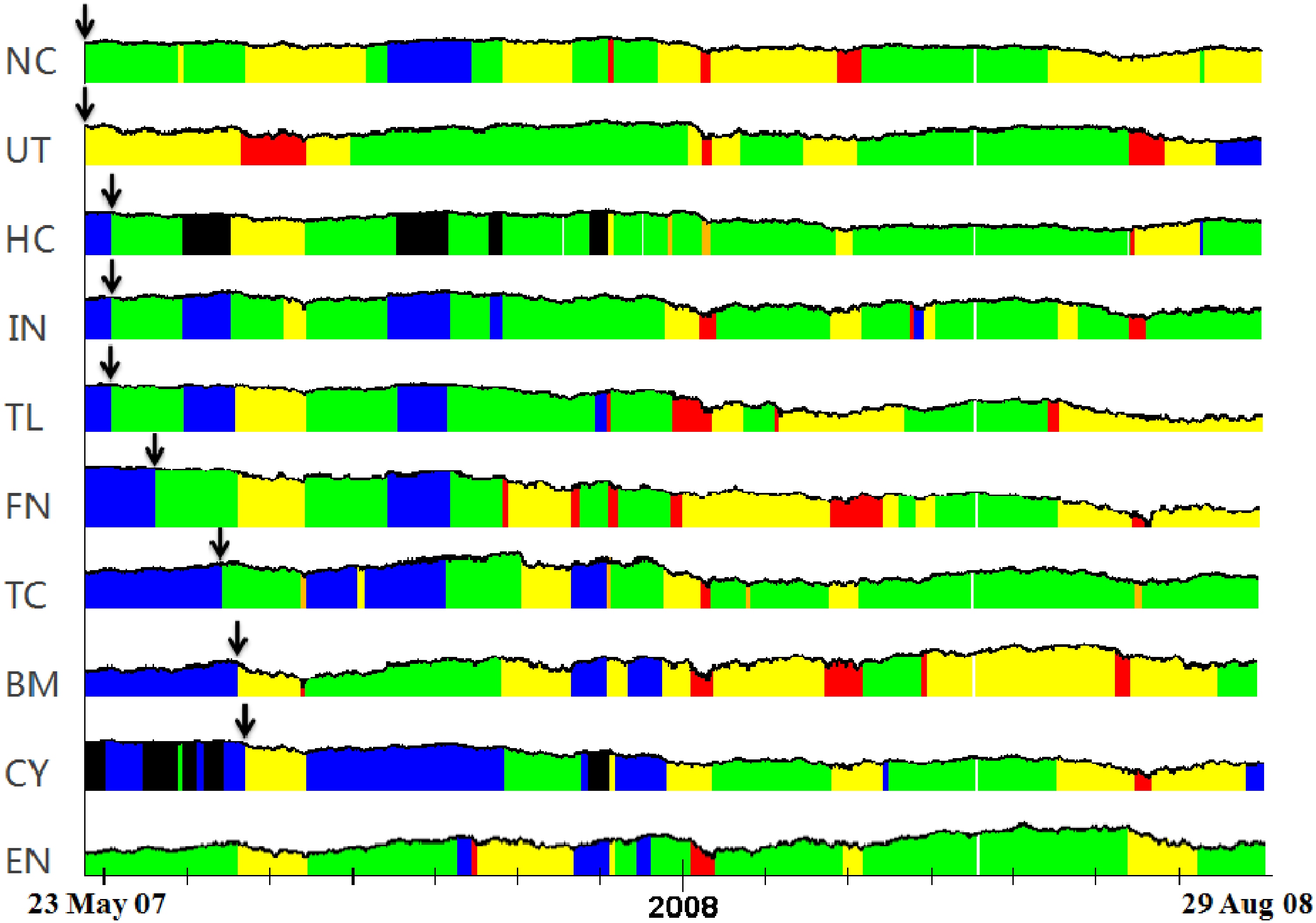
# Time Series Segmentation

| ***k*** | ***Symbol*** | ***Economic Sector*** |
|---|---|---|
| 1 | BM | Basic Materials |
| 2 | CY | Consumer Services |
| 3 | EN | Oil & Gas |
| 4 | FN | Financials |
| 5 | HC | Healthcare |
| 6 | IN | Industrials |
| 7 | NC | Consumer Goods |
| 8 | TC | Technologies |
| 9 | TL | Telecommunications |
| 10 | UT | Utilities |

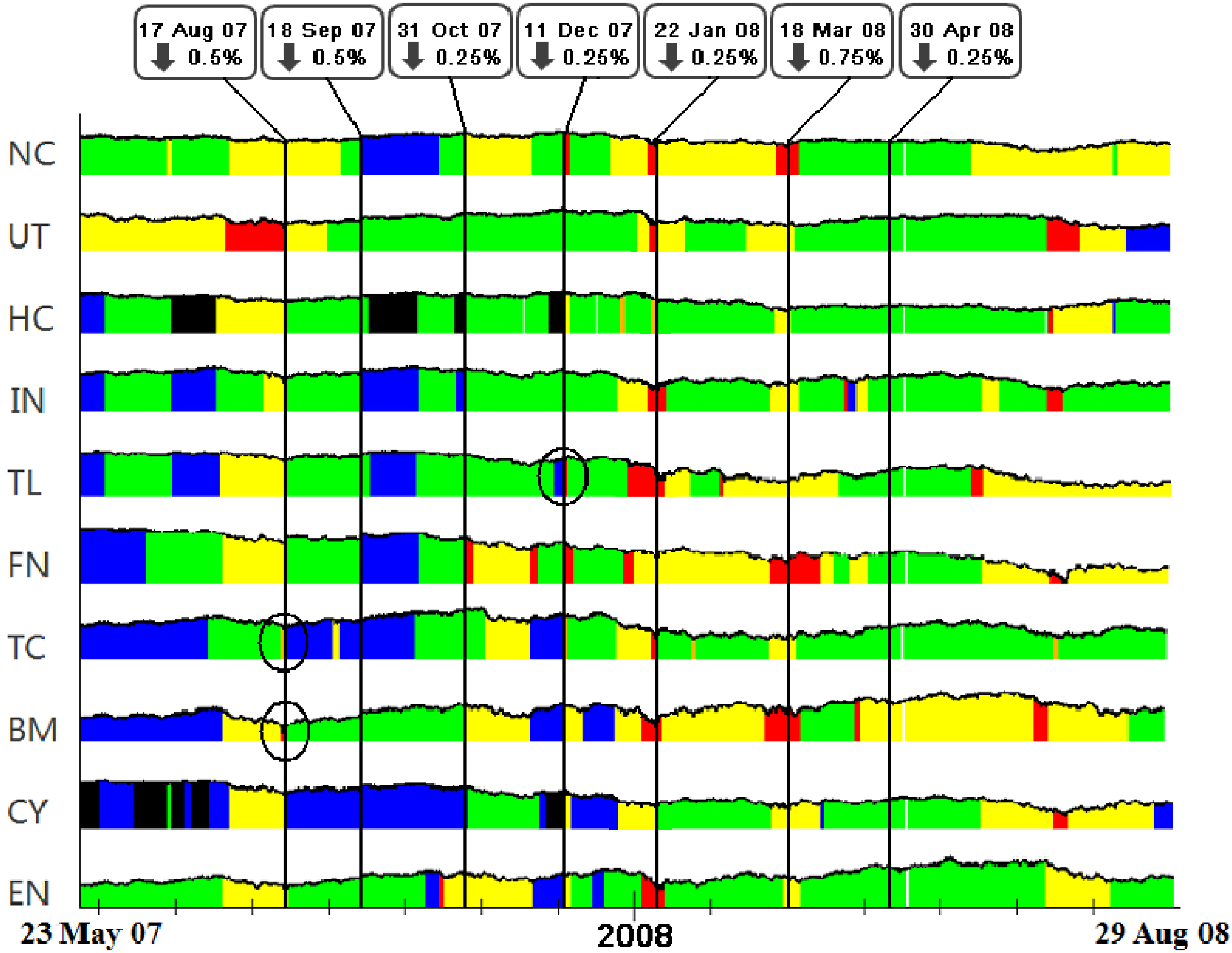
# Time Series Segmentation



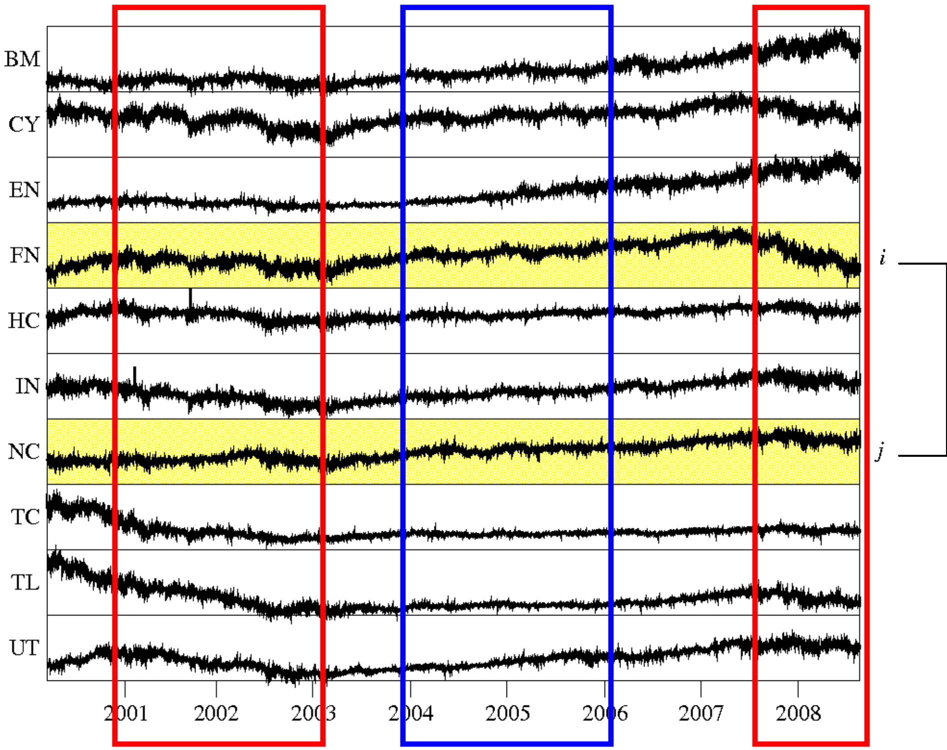
# Time Series Segmentation



# Time Series Segmentation

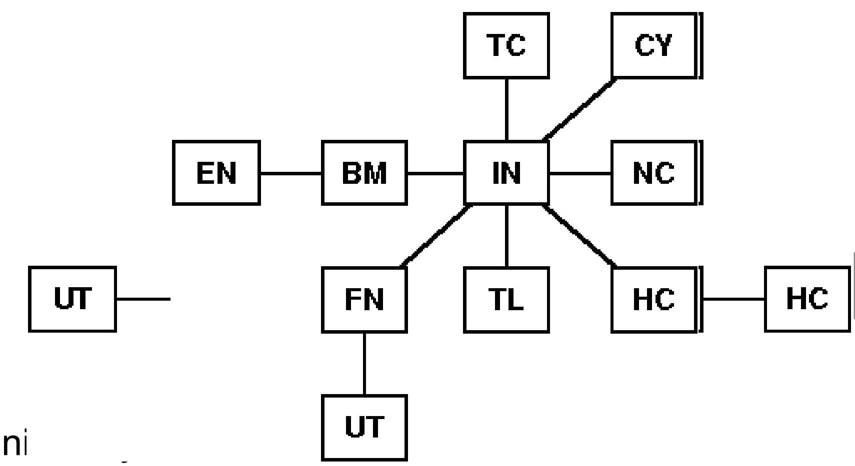


# Time Series Segmentation

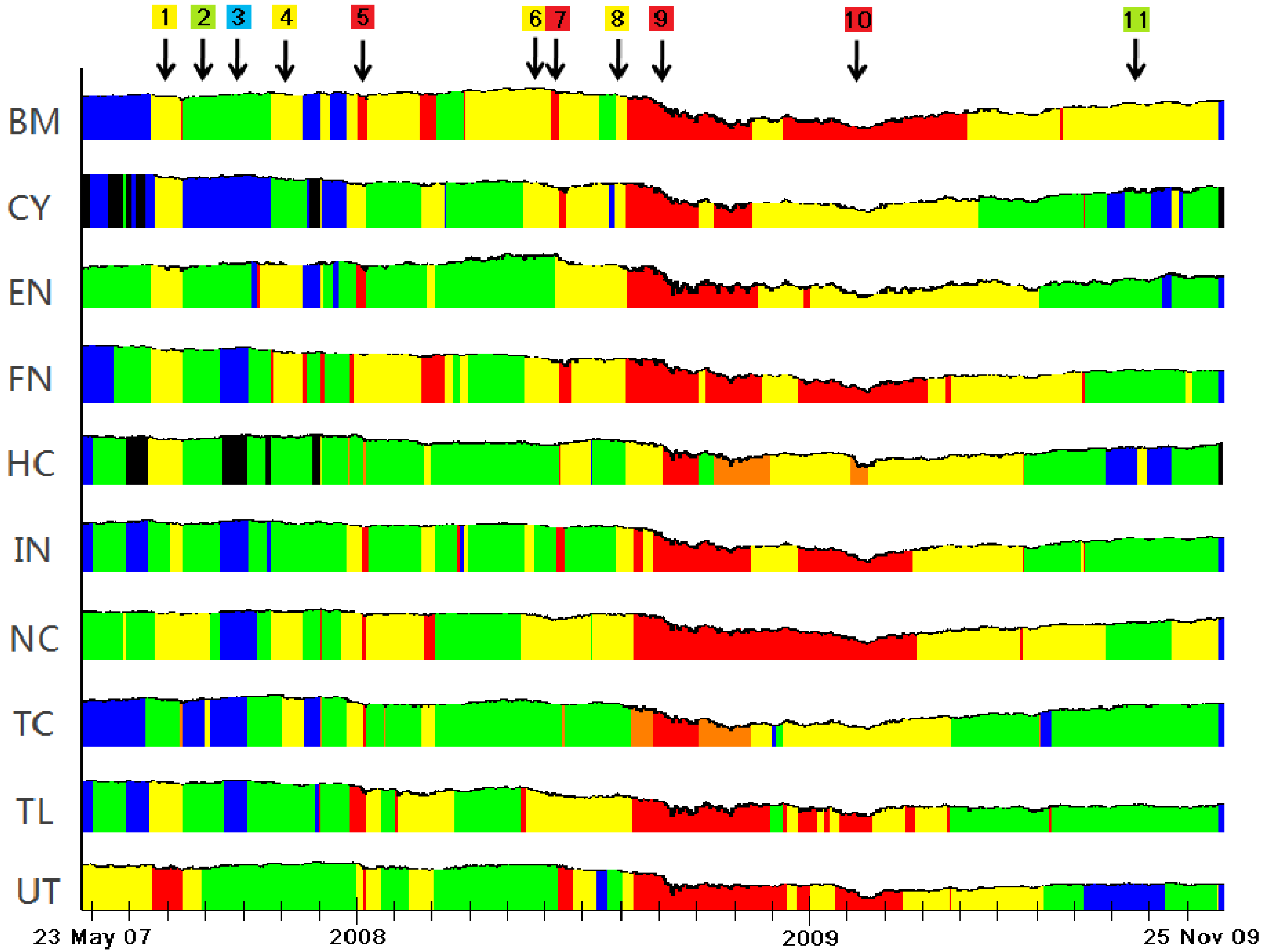



2001-2002 2004-2005 2008-2009

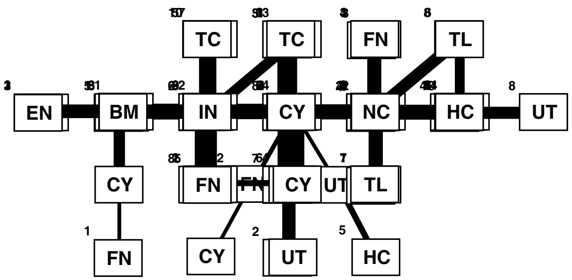
$$C_{ij} = \frac{\left\langle \left(x_i - \bar{x}_i\right)\left(x_j - \bar{x}_j\right)\right\rangle}{\sigma_i \sigma_j} = \left\langle \frac{\delta x_i}{\sigma_i} \frac{\delta x_j}{\sigma_j} \right\rangle$$

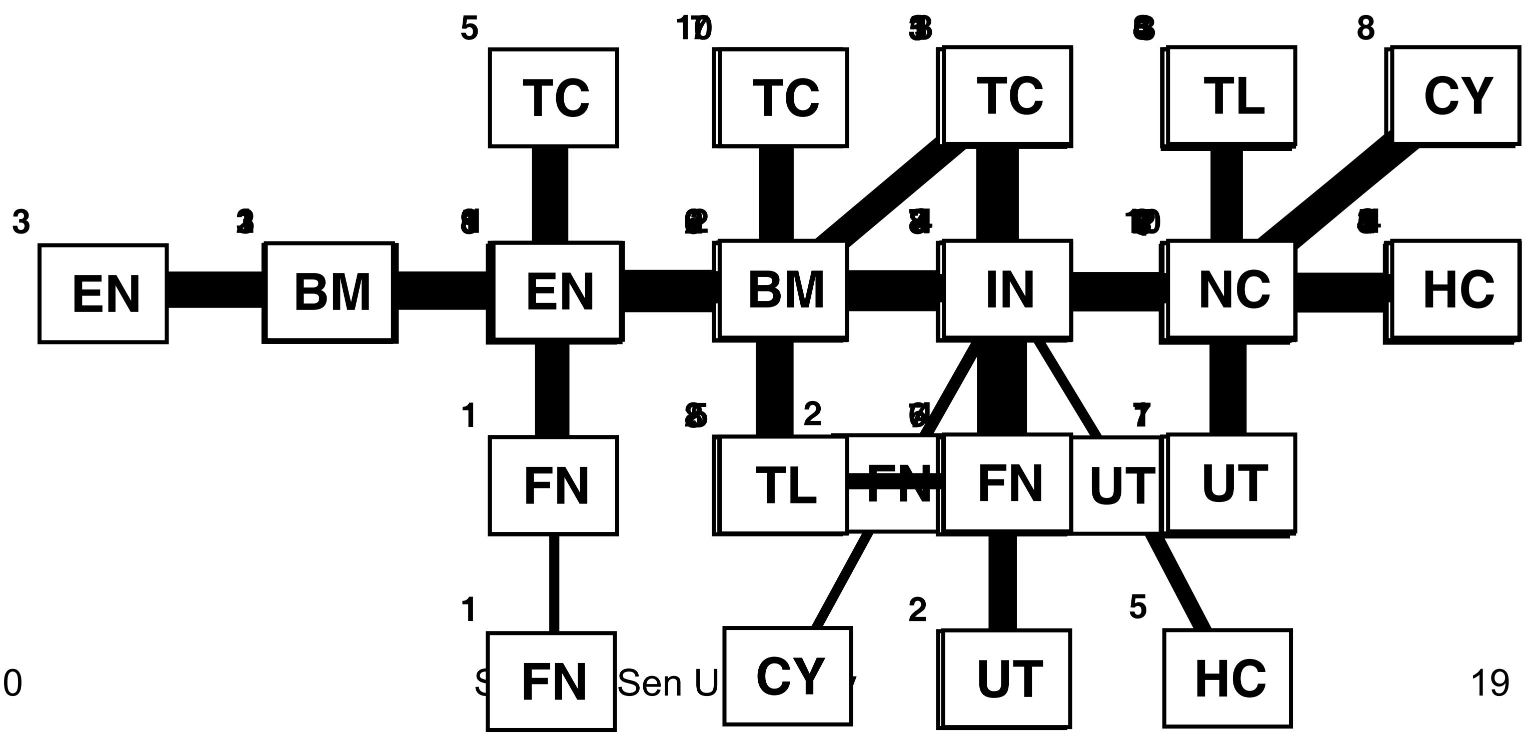


# Time Series Segmentation

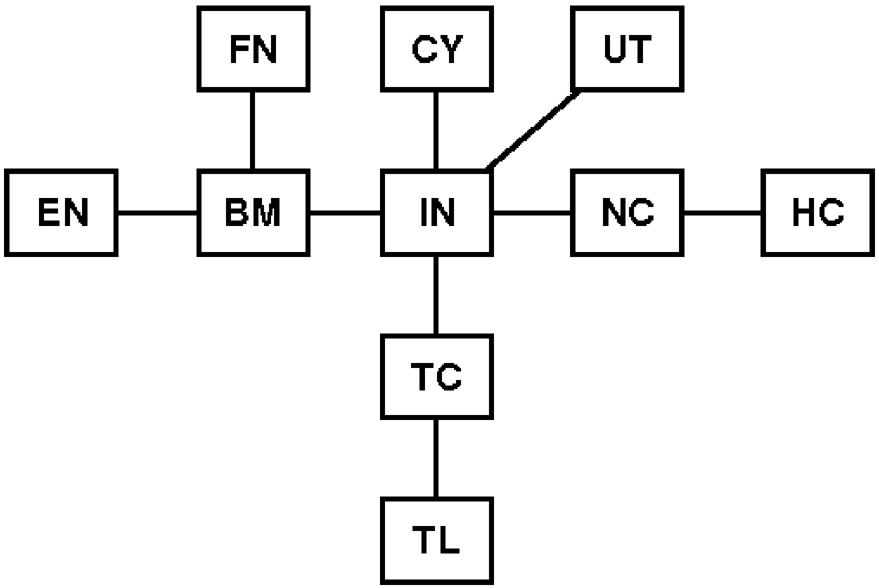


# Time Series Segmentation

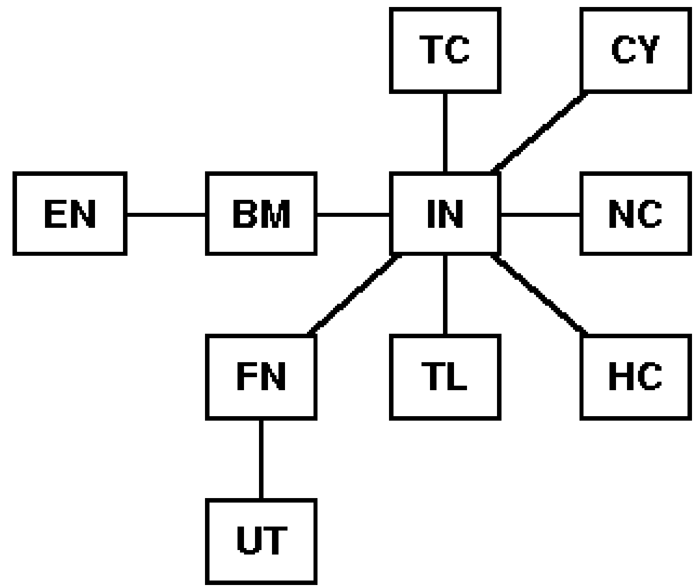


# Time Series Segmentation

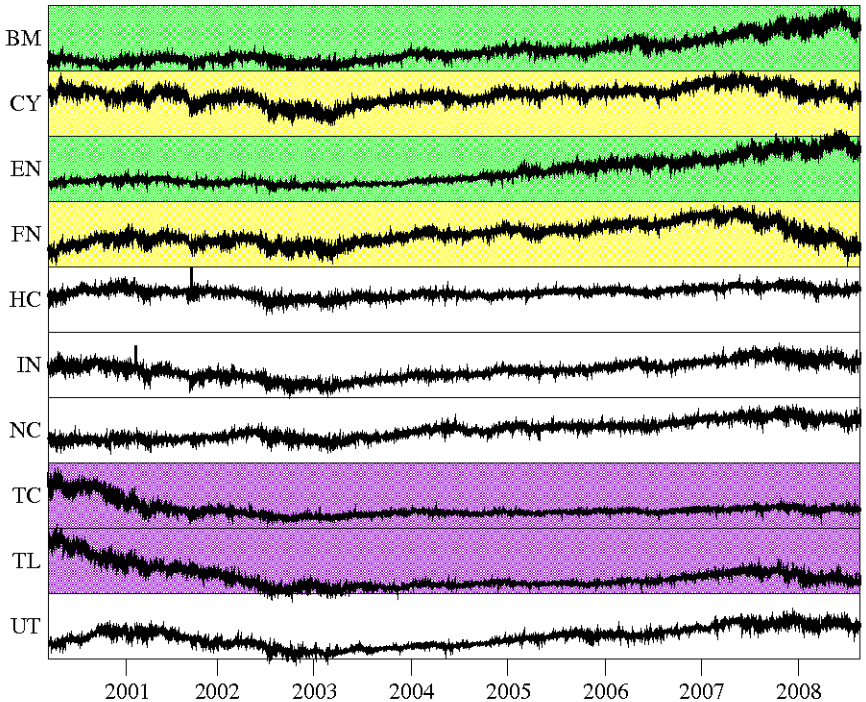



Sep 2009




2004-2005

# Time Series Clustering




$$C_{ij} = \frac{\left\langle (x_i - \bar{x}_i)(x_j - \bar{x}_j) \right\rangle}{\sigma_i \sigma_j} = \left\langle \frac{\delta x_i}{\sigma_i} \frac{\delta x_j}{\sigma_j} \right\rangle$$

$\downarrow$

$$D_{ij} = \left\langle \theta\left( \frac{\delta x_i}{\sigma_i} - 1 \right) \theta\left( \frac{\delta x_j}{\sigma_j} - 1 \right) \right\rangle$$

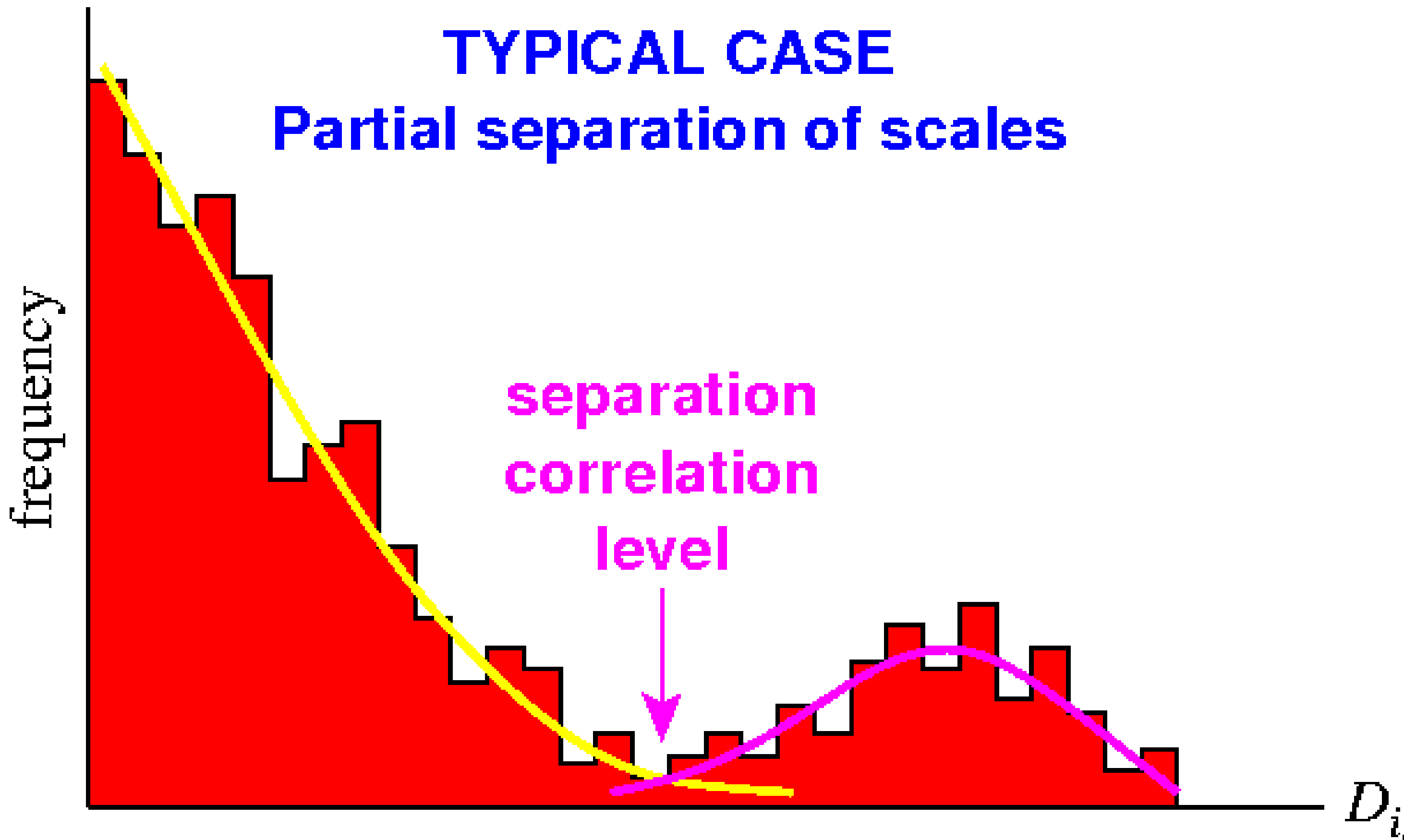
digital cross correlations

$\downarrow$

$$\tilde{D}_{ij} = \left\langle \theta(\Delta x_i \Delta x_j) \right\rangle \text{ or } \tilde{D}_{ij} = \sum_{t=1}^{N} \theta(\Delta x_{it} \Delta x_{jt})$$

comovement digital cross correlations

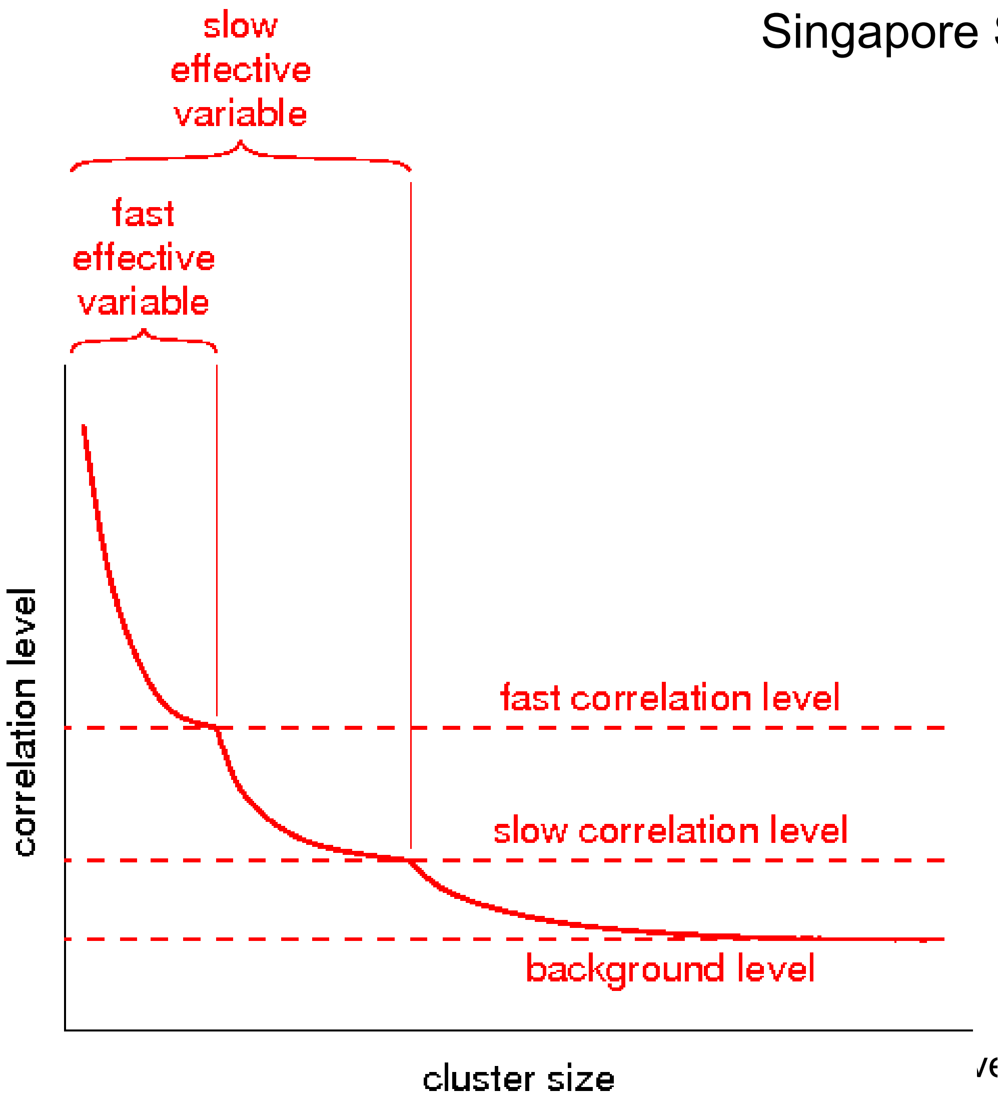
# Time Series Clustering



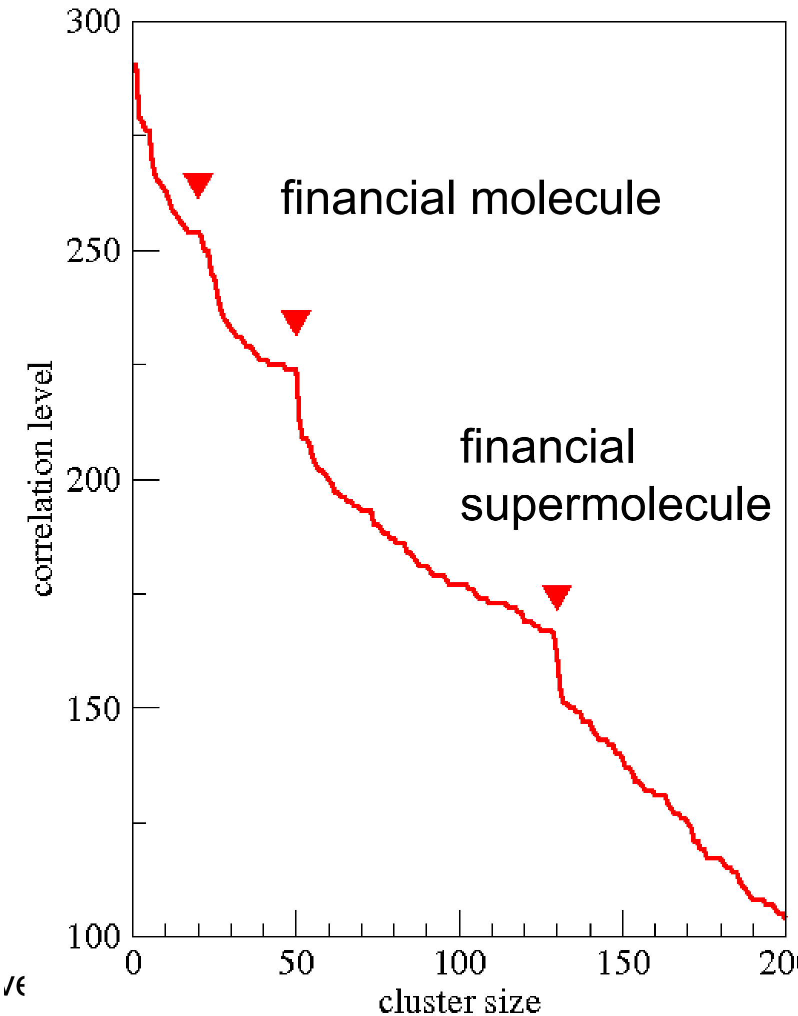
# Time Series Clustering

- Partial Hierarchical Clustering
  - Find $D_{i^*j^*} = \max_{ij} D_{ij}$
  - Use $c = \{i^*, j^*\}$ as seed cluster
  - Add $k^*$ to cluster if
    - $D_{k,c} = \min_{l \in c} D_{kl}$
    - $D_{k^*,c} = \max_k D_{k,c}$
  - Iterate to grow cluster
  - Plot correlation level $D_{k^*,c}$ against cluster size

# Time Series Clustering




Singapore Stock Exchange (SGX): 2006-2007

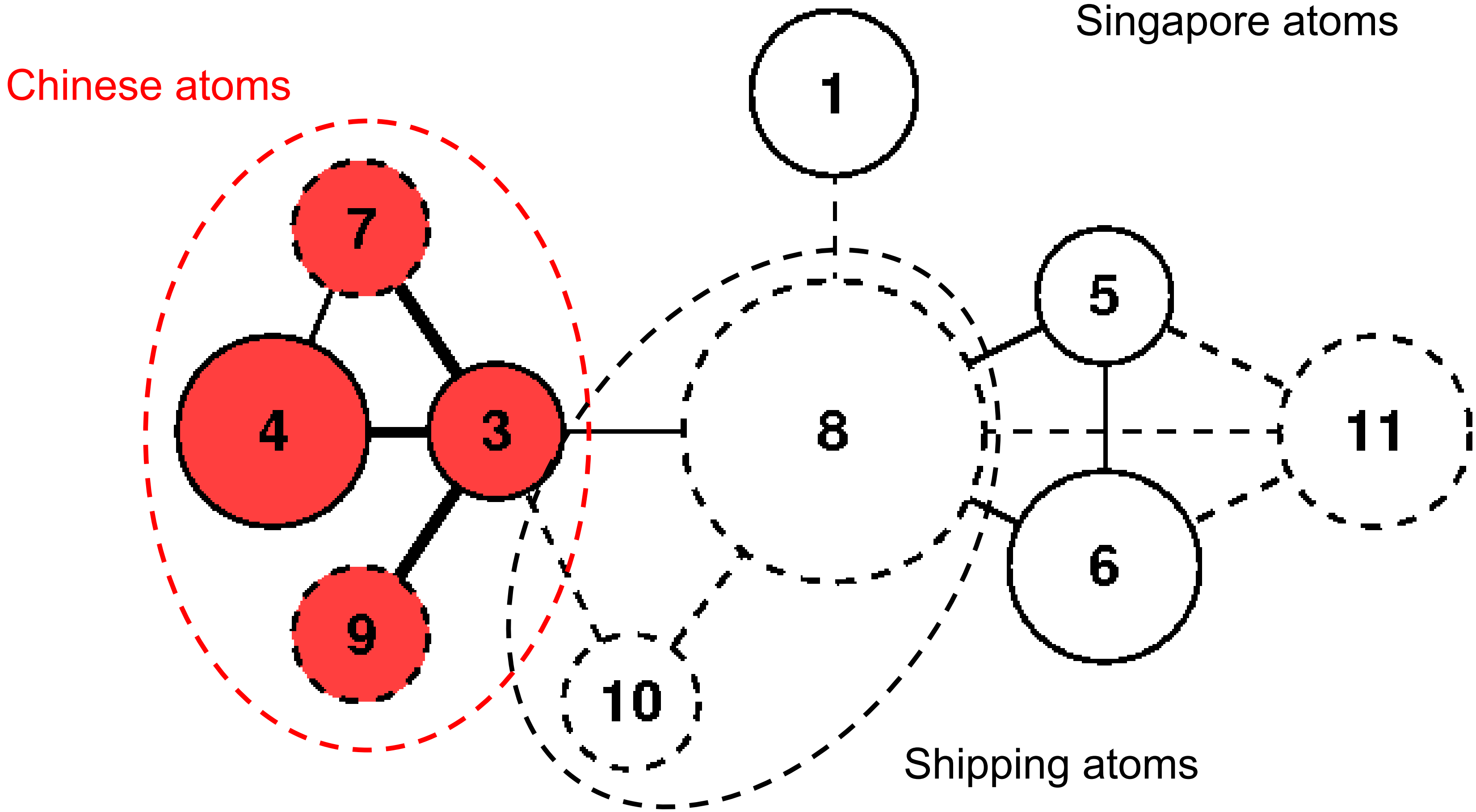


# Time Series Clustering

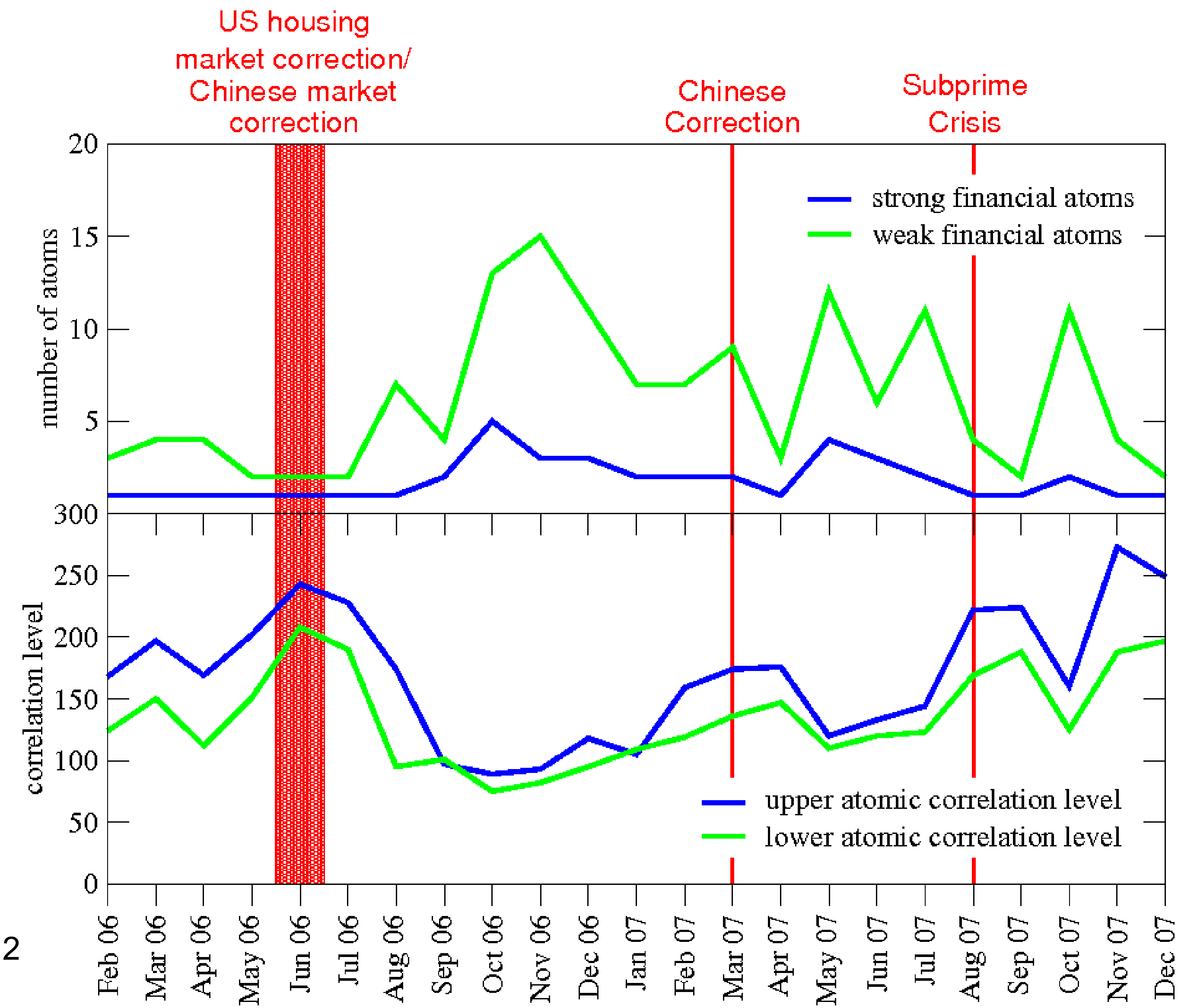
| ***SGX1*** | ***SGX2*** | ***SGX3*** | ***SGX4*** | ***SGX5*** | ***SGX6*** |
|---|---|---|---|---|---|
| Singtel | Singapore Airlines | Celestial Nutri-foods | Mirach Energy | CapitaLand | DBS Gp Hldg |
| Singtel 10 | Singapore Airlines 200 | China Sun Bio-chem Tech Gp | Sky China Petroleum Svcs | City Develop-ment | United Overseas Bank |
| Singtel 100 | | | Ferrochina | | Overseas Chinese Banking Corp |
| | | | China Sky Chemical Fibre Co | | Wing Tai Hldg |

# Time Series Clustering

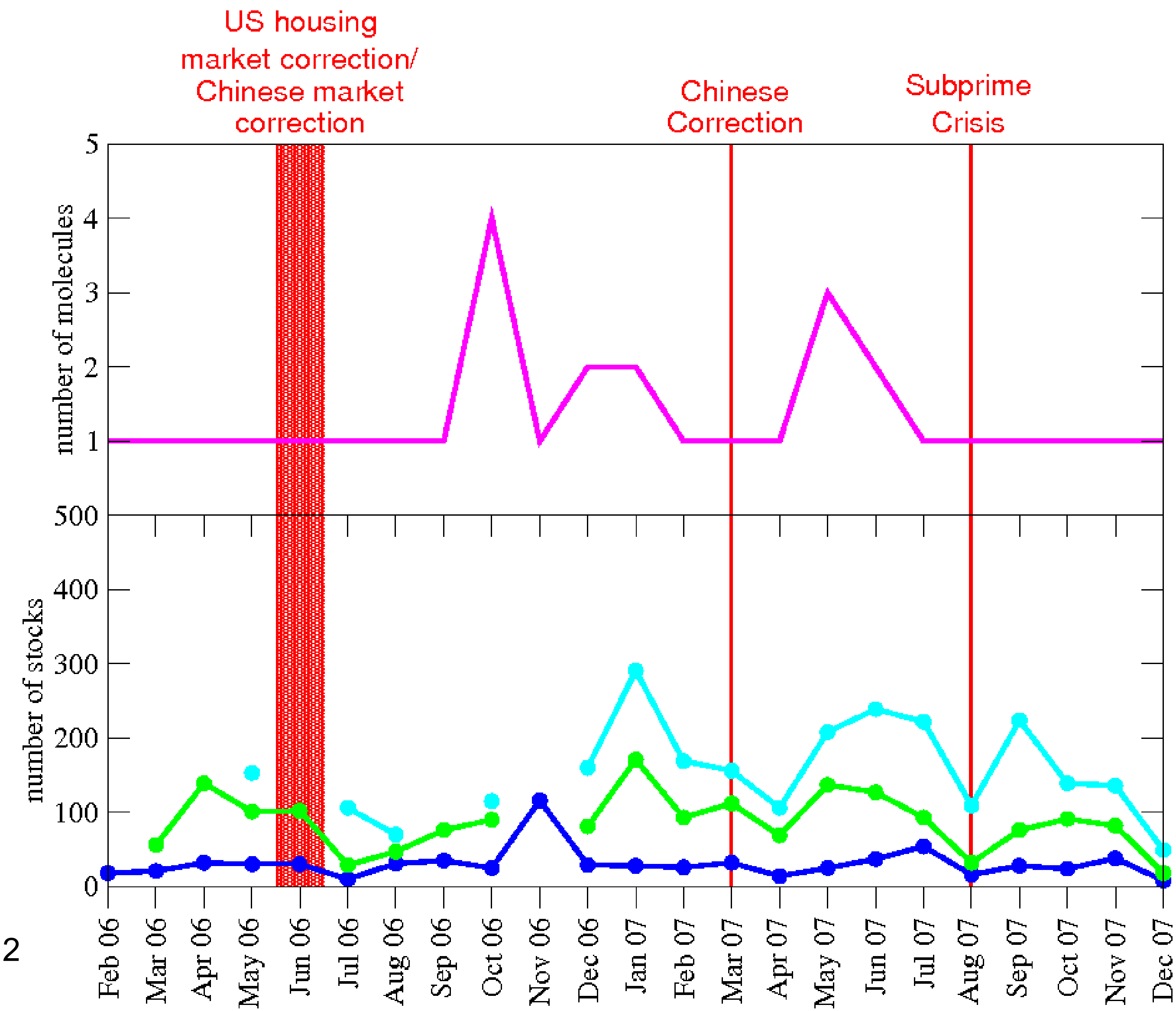
SGX financial molecule 2006-2007



# Time Series Clustering



# Time Series Clustering



# Conclusions

- Possible to identify macroscopic phases and study their dynamics from microscopic time series
- Time series segmentation
  - US economy
    - Crisis and growth
    - MST topologies
- Time series clustering
  - SGX
    - Financial atoms and molecules
    - Chemical picture of market crashes

# Acknowledgments

- Time Series Segmentation
  - WONG Jian Cheng
  - Gladys LEE Hui Ting
  - ZHANG Yiting
  - KOK Jun Liang
  - Dr Manamohan PRUSTY
- Time Series Clustering
  - GOO Yik Wen
  - LIAN Tong Wei
  - ONG Wei Guang
  - CHOI Wen Ting
  - Mridula DAMODARAN

*Thank You!*