

ТЕОРИЯ ВЕРОЯТНОСТЕЙ, СЛУЧАЙНЫЕ ПРОЦЕССЫ,  
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА  
И ПРИЛОЖЕНИЯ

**Петерис Грабуст**

Резекненская Высшая школа  
г. Резекне, Латвия

**Способы оценок сходства  
временных рядов**

Белорусский государственный университет, 22 – 25 февраля 2010 г.

# Постановка задачи

**Временный ряд** – формально это последовательность данных во временном интервале.

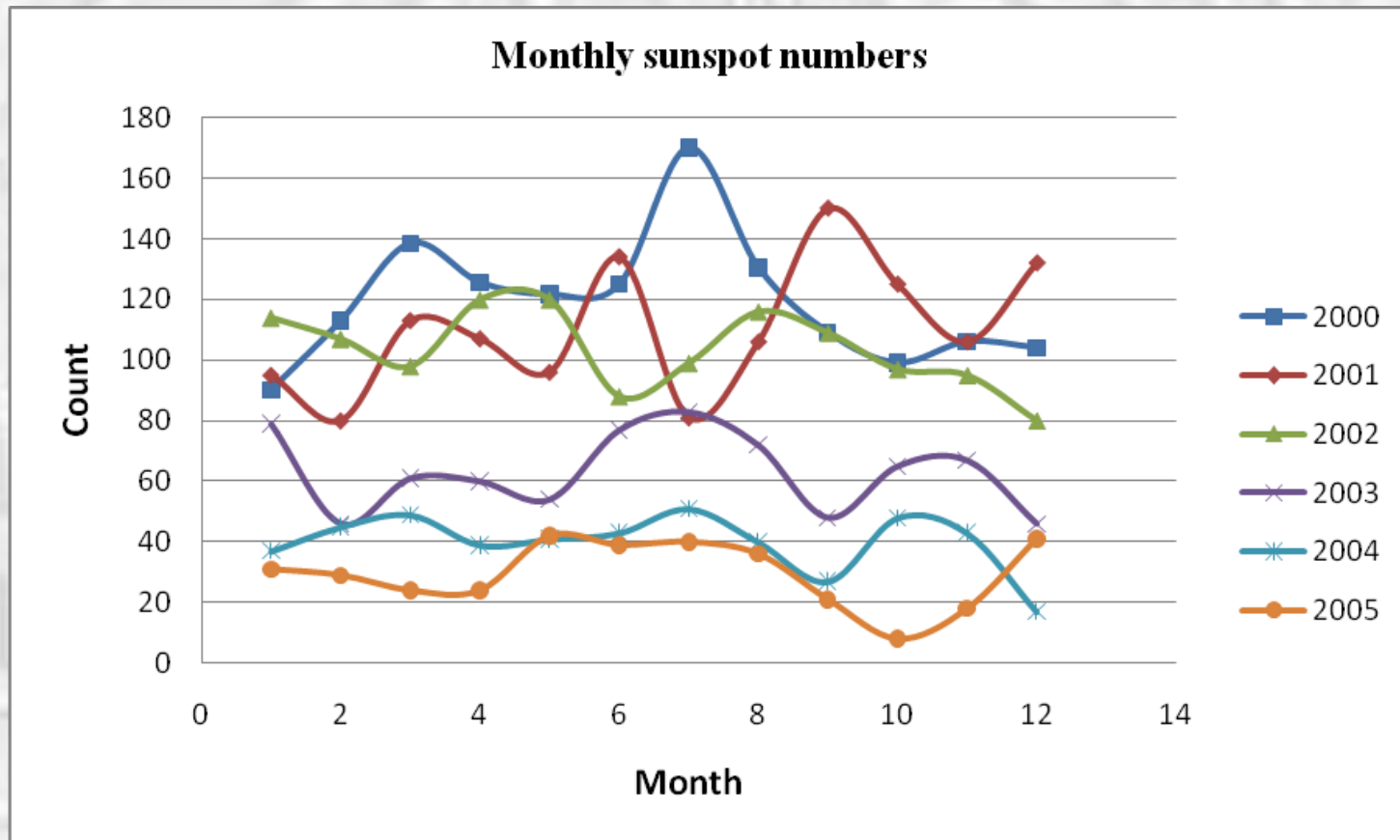
**Анализ временных рядов** достаточно хорошо известная задача, но в последнее время производятся исследования с целью использовать кластеризацию для анализа временных рядов.

**Главная мотивация** для отображения временных рядов в кластерном виде – это лучшее представление основных характеристик временных рядов.

**Целью данной работы** является исследование методологии определения возможных способов оценки сходства или близости временных рядов и применение этих оценок для анализа результатов кластеризации.

# Анализ временных рядов

- Working Group of Surface Pressure - [http://www.cdc.noaa.gov/gcos\\_wgsp/Timeseries/SUNSPOT/](http://www.cdc.noaa.gov/gcos_wgsp/Timeseries/SUNSPOT/)



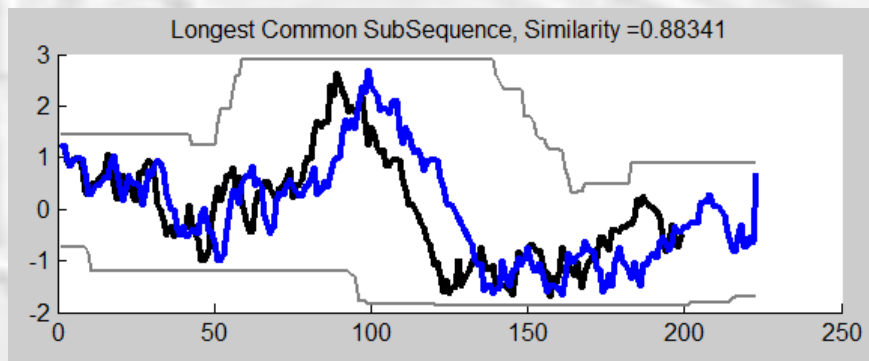
# Меры сходства

- Классическая Евклидова дистанция

$$D(Q,S) = \sqrt{\sum_{i=1}^n (q_i - s_i)^2}$$

- Метод наиболее длинной общей последовательности (LCSS)

$$LCSS[i, j] = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0 \\ 1 + LCSS[i - 1, j - 1] & \text{if } a_i = b_i \\ \max(LCSS[i - 1, j], LCSS[i, j - 1]) & \text{otherwise} \end{cases}$$

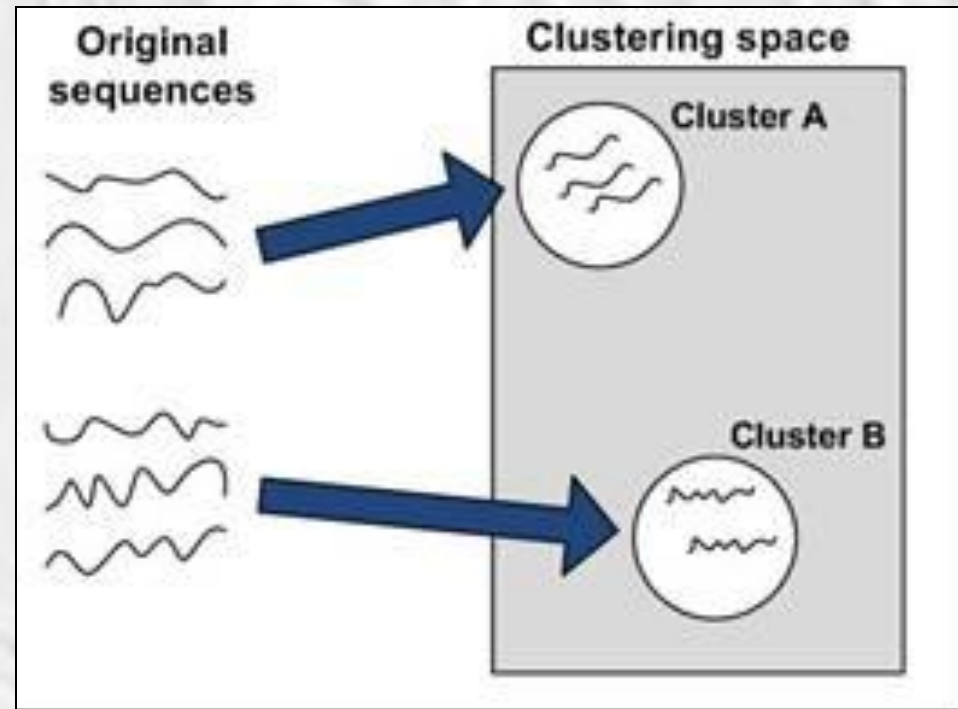


# Альтернатива

- Кластеризация (K-means, Hierarchical)

## K-means clustering

1. Decide on a value for  $k$ .
2. Initialize the  $k$  cluster centers (randomly, if necessary).
3. Decide the class memberships of the  $N$  objects by assigning them to the nearest cluster center.
4. Re-estimate the  $k$  cluster centers, by assuming the memberships found above are correct.
5. If none of the  $N$  objects changed membership in the last iteration, exit. Otherwise goto 3.

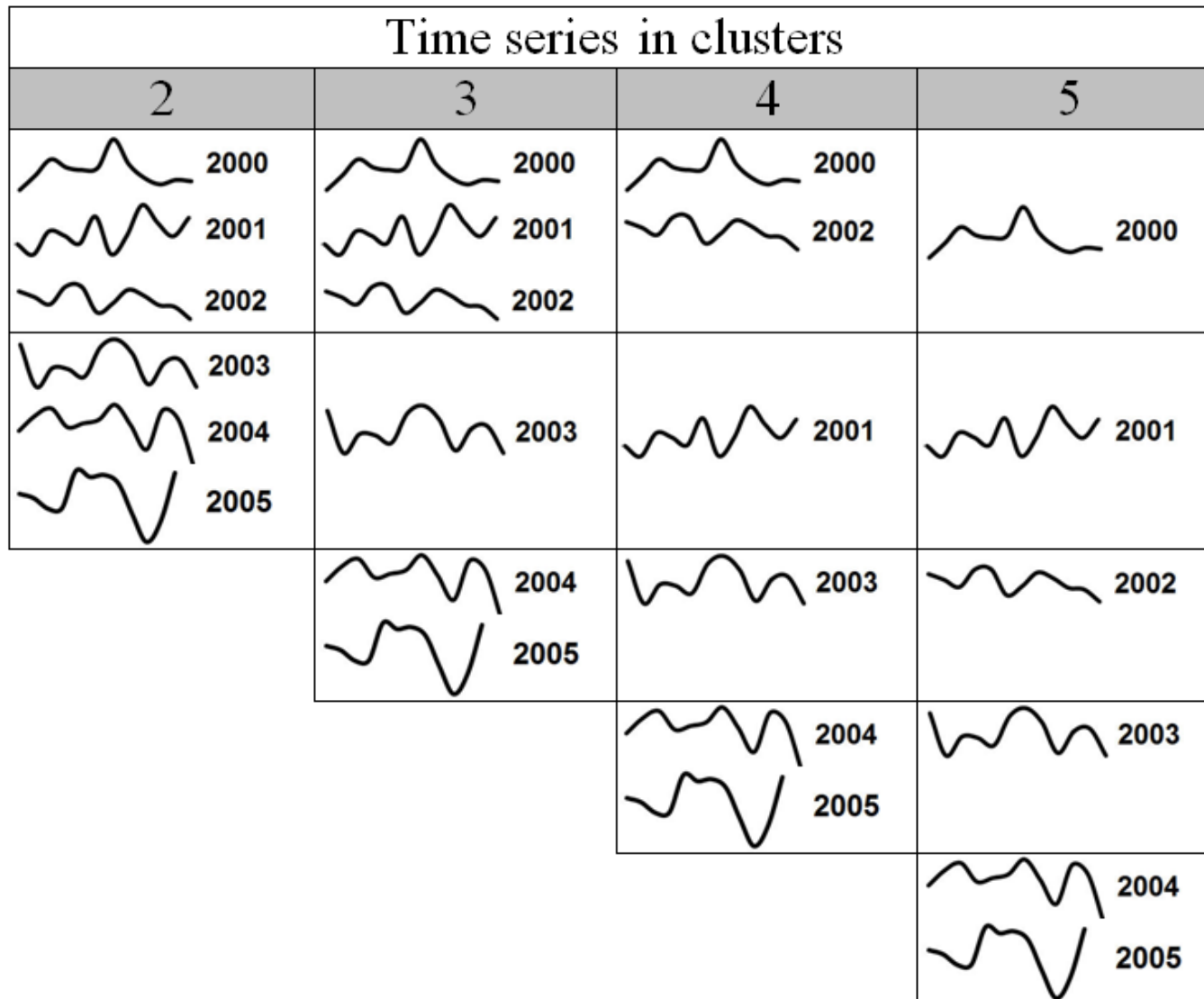


# Применение LCSS

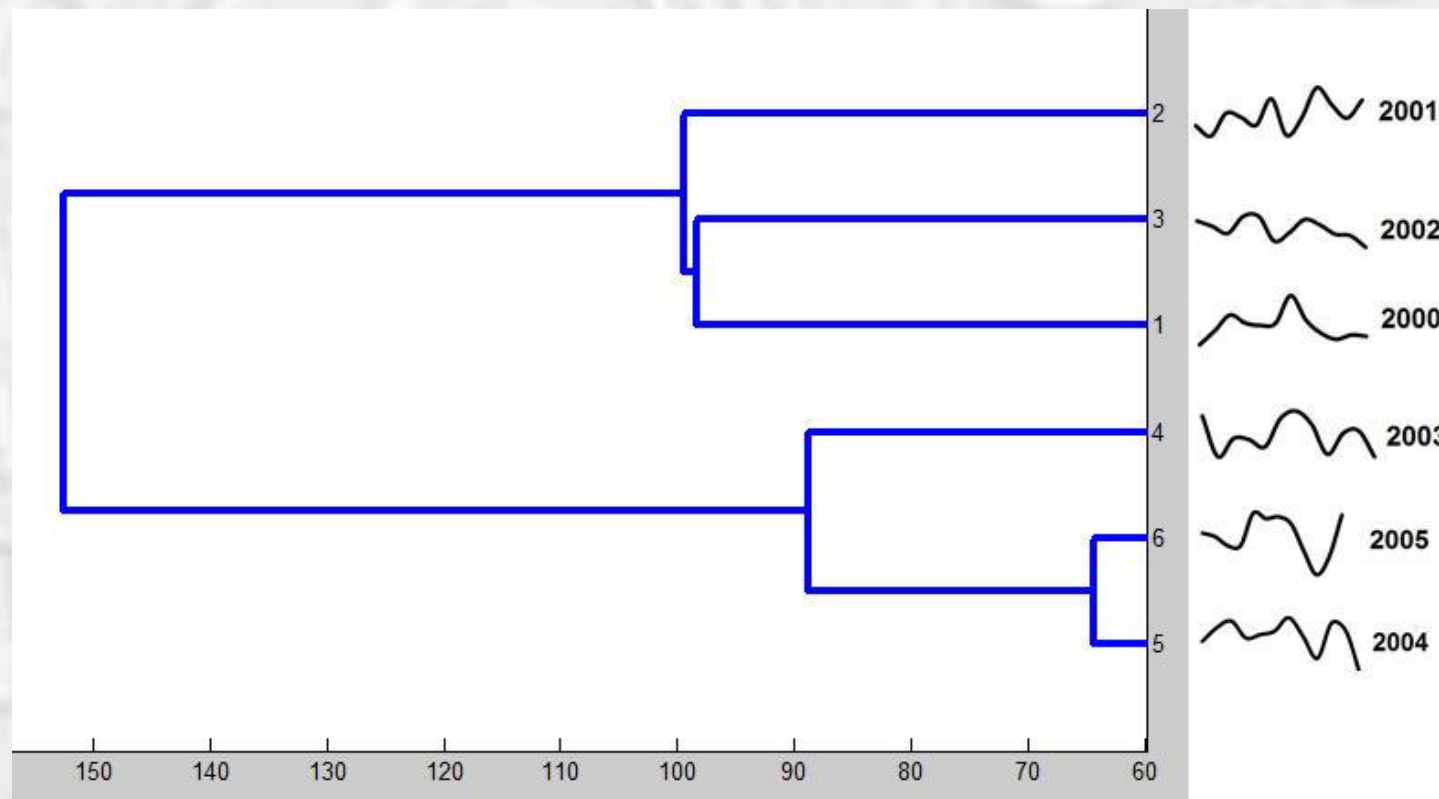
Сравнение временных рядов попарно

	2000	2001	2002	2003	2004	2005
2000	1	0.17	0.08			
2001	0.17	1				
2002		0.08	1			
2003				1	0.16	0
2004				0.16	1	0.25
2005					0.25	1

# Применение K-means кластеризации



# Применение Hierarchical кластеризации





# Выводы

По итогам экспериментов сделано заключение, что результаты кластеризации конкретного набора временных рядов соответствует полученным результатам с помощью метода LCSS - т.е. **результаты адекватны.**

**Спасибо за внимание !**