Data Analytics

Course: 18-787

Patrick McSharry
patrick@mcsharry.net
www.mcsharry.net
Twitter: @patrickmcsharry

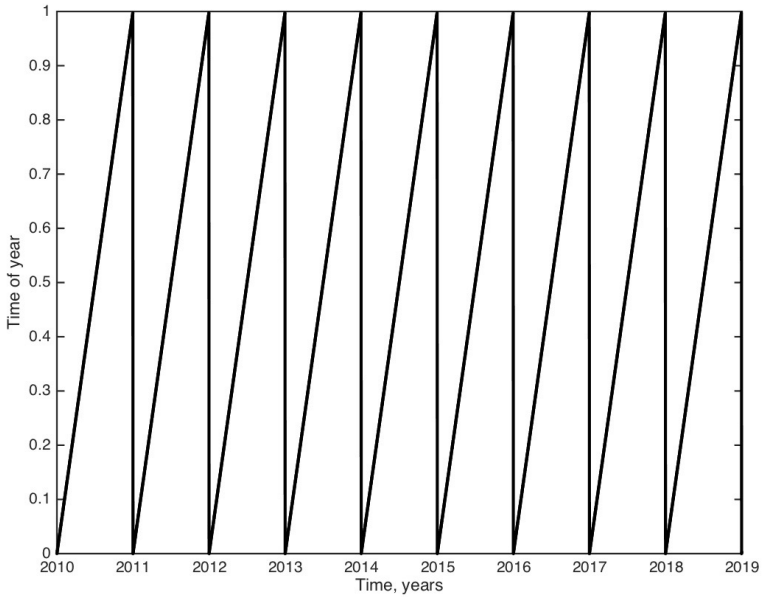
Spring 2023

ICT Center of Excellence
Carnegie Mellon University

Data Analytics

WEEK 6A

A3Q3



- Time of year variable serves to wrap time around a circle. For several years of data, this helps to detect intra-annual seasonality and build models (see wind power and electricity demand forecasting). Modulus operator is required to reset to 0 at the start of each year. Visually this is easy but the code for doing it properly is best tested with several years of data:
- t = datenum('01-Jan-2010'):datenum('31-Dec-2018');

A3Q8

- Is there a statistically significant difference between demand during the weekend (Sat & Sun) and during the working week (Mon-Fri)?
- Create two distributions in vectors: X = demand values during Sat & Sun and Y = demand values Mon-Fri.
- H = ttest2(X,Y) performs a t-test of the hypothesis that two independent samples, in the vectors X and Y, come from distributions with equal means, and returns the result of the test in H.
- H=0 indicates that the null hypothesis ("means are equal") cannot be rejected at the 5% significance level. H=1 indicates that the null hypothesis can be rejected at the 5% level.

https://www.mathworks.com/help/stats/ttest2.html

Course outline

Week	Lecture A	Lecture B
1	Data Analytics	Weather forecasting
2	Renewable energy	Wind energy
3	Solar energy	Demand forecasting
4	Risk	Extreme events
5	Health	Biomedicine
6	Early warning systems	Economic forecasting

Today’s Lecture

No.	Activity	Description	Time
1	Challenge	Preventing disaster	10
2	Discussion	Components of EWS	10
3	Case study	Parkinson’s disease	10
4	Analysis	Nuclear reactor clogging	20
5	Demo	UK Met Office	20
6	Q&A	Questions and feedback	10

Hazards, vulnerability and disaster

- Natural hazards, such as storms, droughts, volcanic eruptions, or earthquakes, need not spell disaster.
- A disaster occurs only if a community or population is exposed to the natural hazard and cannot cope with its effects.
- Catastrophe equation:

  Risk = Hazard x Exposure x Vulnerability

Examples of vulnerability

- Torrential rain in the middle of an ocean will not cause a disaster, but the same heavy rainfall on a vulnerable population – say a town on the side of a hillside stripped of trees - may result in landslides and a huge loss of life.
- A minor drought may cause a famine if a region's agricultural production is highly stressed by civil war.
- A community that lacks an early warning system may sleep while volcanic ash clouds bear down upon them.
- Vulnerability is the potent additive that mixes with natural hazards to cause disasters.

Early Warning Systems

- Early Warning Systems (EWS) were first developed for addressing the threats posed by natural disasters.
- This approach is now being applied to many other areas.
- The name **early** warning implies that the warning must provide adequate notice that stakeholders can take action so as to decrease any adverse impacts.

Poll: Early warning system

- What are the key ingredients of an early warning system?

- **Slido.com #66086**

EWS components

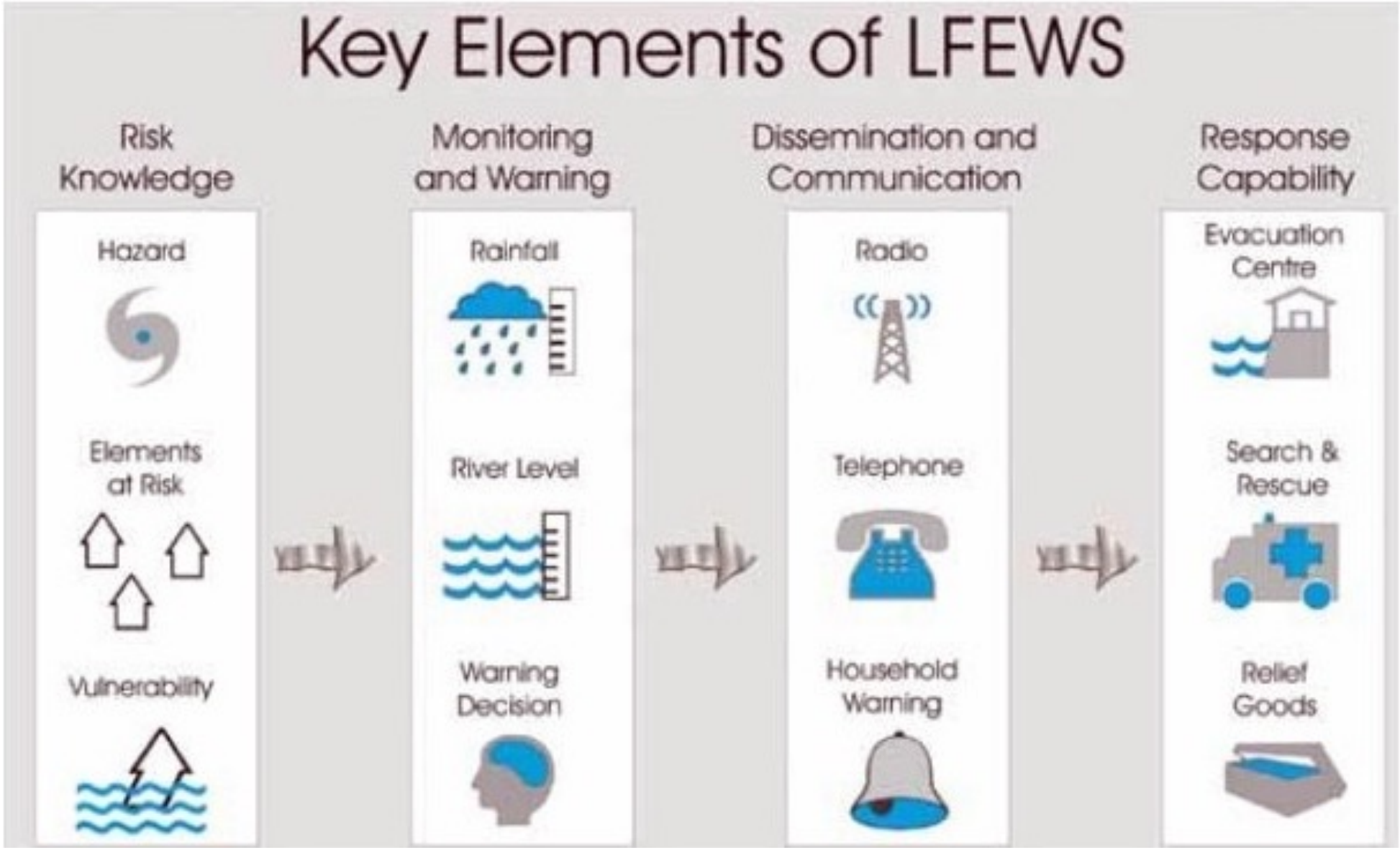
- There are four important components required to construct an EWS:
  - risk analysis;
  - monitoring and warning;
  - dissemination and communication;
  - response capability.
- Warnings must be timely, accurate, unambiguous and credible.

Effective EWS

Risk Analysis	Monitoring and Warning	Dissemination and Communication	Response Capability
Systematically collect data and undertake risk assessments	*Develop hazard monitoring and early warning services*	*Communicate risk information and early warnings*	*Build national and community response capabilities*
Are the hazards and the vulnerabilities well known? What are the patterns and trends in these factors? Are risk maps and data widely available?	Are the right parameters being monitored? Is there a sound scientific basis for making forecasts? Can accurate and timely warnings be generated?	Do warnings reach all of those at risk? Are the risks and the warnings understood? Is the warning information clear and useable?	Are response plans up to date and tested? Are local capacities and knowledge made use of? Are people prepared and ready to react to warnings?

Source: www.unisdr.org

Flood EWS



Source: http://www.planet-action.org

EWS in operation

- The major players concerned with the different elements meet regularly to ensure they understand all of the other components and what other parties need from them.
- Risk scenarios are constructed and reviewed.
- Specific responsibilities throughout the chain are agreed and implemented.

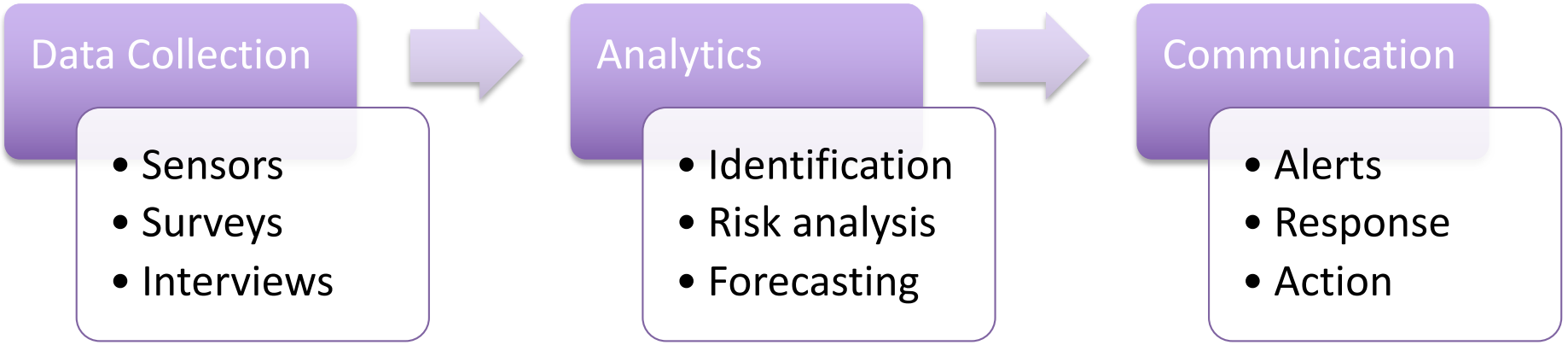
EWS in operation contd.

- Past events are studied and improvements are made to the early warning system.
- Manuals and procedures are agreed and published.
- Communities are consulted and information is disseminated.
- Operational procedures such as evacuations are practiced and tested.

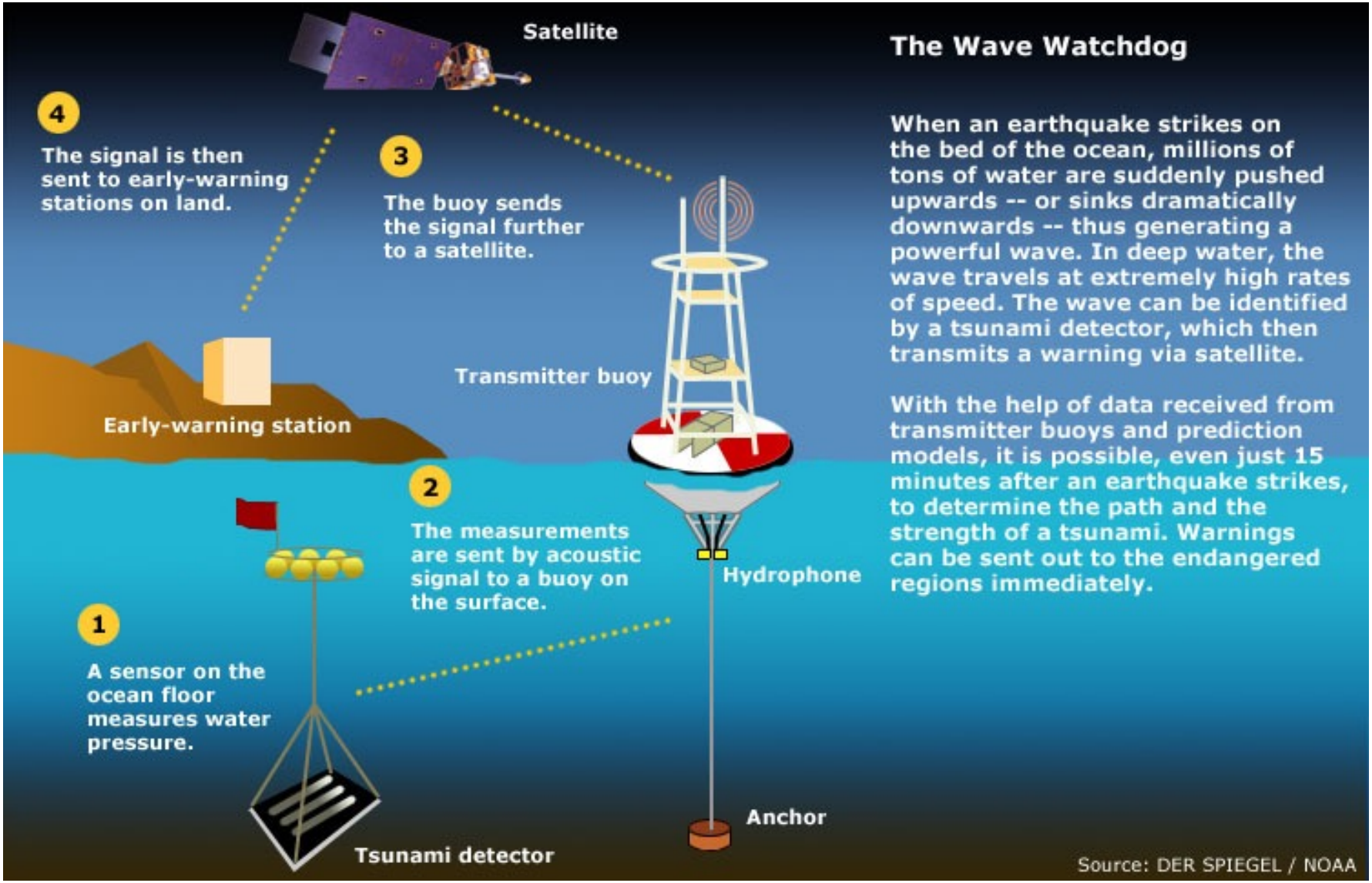
Data Analytics and EWS

- Data analytics and quantitative modelling can be used to provide warnings and actionable alerts about extreme events.
- Sensors, mobiles and wireless connectivity means that it is easier to collect and centralize data storage.
- Algorithms are used to assess risk
- Mobiles can communicate risk levels and suggest appropriate action.

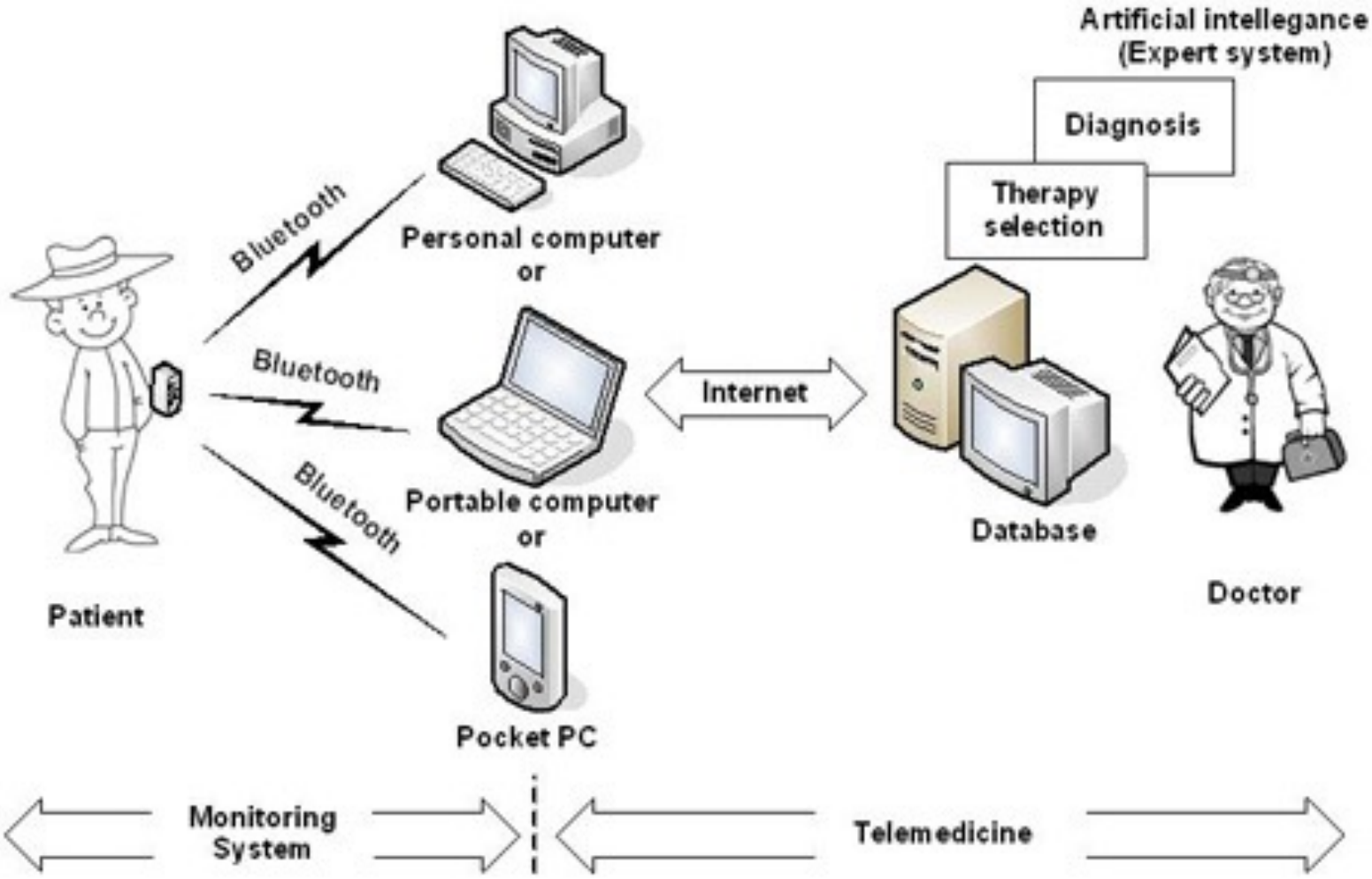
EWS process


Data Collection
• Sensors
• Surveys
• Interviews
Analytics
• Identification
• Risk analysis
• Forecasting
Communication
• Alerts
• Response
• Action

Tsunami EWS


Satellite
4
The signal is then sent to early-warning stations on land.
3
The buoy sends the signal further to a satellite.
Transmitter buoy
Early-warning station
2
The measurements are sent by acoustic signal to a buoy on the surface.
Hydrophone
1
A sensor on the ocean floor measures water pressure.
Anchor
Tsunami detector
The Wave Watchdog
When an earthquake strikes on the bed of the ocean, millions of tons of water are suddenly pushed upwards -- or sinks dramatically downwards -- thus generating a powerful wave. In deep water, the wave travels at extremely high rates of speed. The wave can be identified by a tsunami detector, which then transmits a warning via satellite.
With the help of data received from transmitter buoys and prediction models, it is possible, even just 15 minutes after an earthquake strikes, to determine the path and the strength of a tsunami. Warnings can be sent out to the endangered regions immediately.
Source: DER SPIEGEL / NOAA

Medical EWS



Source: www.medgadget.com

EWS for critical illness

MEWS (Modified Early Warning System)							
	3	2	1	0	1	2	3
Respiratory Rate per minute		Less than 8		9-14	15-20	21-29	More than 30
Heart Rate per minute		Less than 40	40-50	51-100	101-110	111-129	More than 129
Systolic Blood Pressure	Less than 70	71-80	81-100	101-199		More than 200	
Conscious level (AVPU)	**U**nresponsive	Responds to **P**ain	Responds to **V**oice	**A**lert	New agitation Confusion		
Temperature (°c)		Less than 35.0	35.1-36	36.1-38	38.1-38.5	More than 38.6	
Hourly Urine For 2 hours	Less than 10mls / hr	Less than 30mls / hr	Less than 45mls / hr				

EARLY WARNING SCORING SYSTEM FOR DETECTING ADULT PATIENTS WHO HAVE OR ARE DEVELOPING CRITICAL ILLNESS

IS THE SCORE FOR YOUR PATIENT 1-2? PERFORM 2 HOURLY OBSERVATIONS AND INFORM NURSE IN CHARGE

IS THE SCORE FOR YOUR PATIENT 3? PERFORM 1-2 HOURLY OBSERVATIONS AND INFORM NURSE IN CHARGE

*IF THE MEWS SCORE IS DETERIORATING : THE WARD S.H.O. OR DUTY DOCTOR **MUST** ATTEND*

IS THE SCORE FOR YOUR PATIENT 4 OR MORE? PERFORM OBSERVATIONS AT LEAST 1/2 HOURLY. ENSURE MEDICAL ADVICE IS SOUGHT AND CONTACT OUTREACH TEAM (see below)

APGAR score

- The Apgar score is the first test given to a newborn, usually in the delivery room.
- The test was designed to quickly evaluate a newborn's physical condition and to see if there's an immediate need for extra medical or emergency care.
- Developed in 1952 by an anesthesiologist named Virginia Apgar.
- Acronym for: Appearance, Pulse, Grimace, Activity, and Respiration.

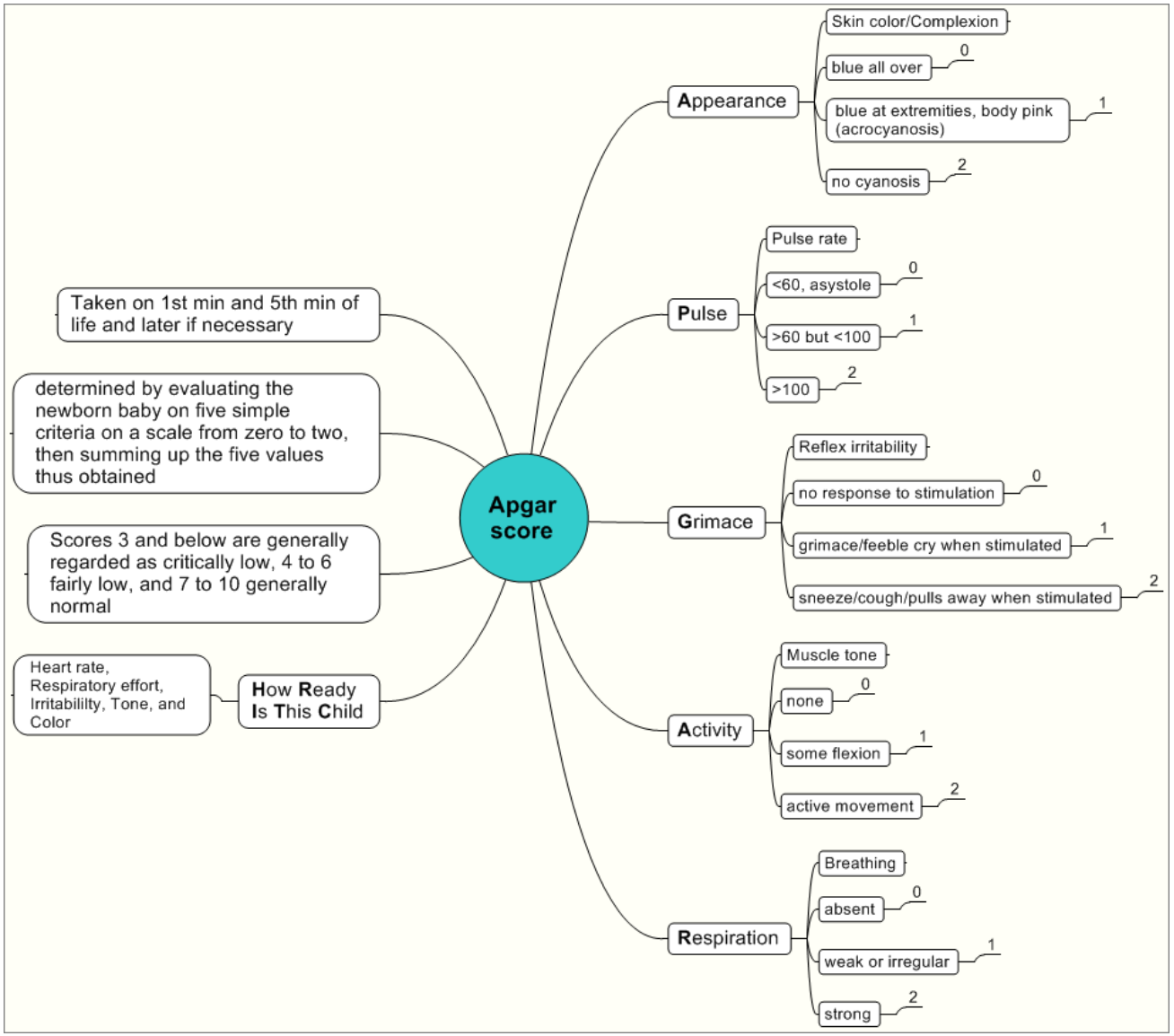
Five criteria of Apgar score

Component\Score	Score of 0	Score of 1	Score of 2
Appearance	Blue or pale all over	Blue at extremities Body pink	Body and extremities pink
Pulse	Absent	<100	>100
Grimace	No response to stimulation	Grimace on suction or aggressive stimulation	Cry on stimulation
Activity	None	Some flexion	Flexed arms and legs that resist extension
Respiration	Absent	Weak, irregular, gasping	Strong, lusty, cry

Interpretation of Apgar

Apgar Score	Interpretation	Response to score after 10-30 minutes
1	Critically Low	Risk of neurological damage and small increase in risk of cerebral palsy
2		
3		
4	Fairly Low	Baby requires medical attention
5		
6		
7	Normal	No medical treatment required
8		
9		
10		

Apgar Score Mind Map

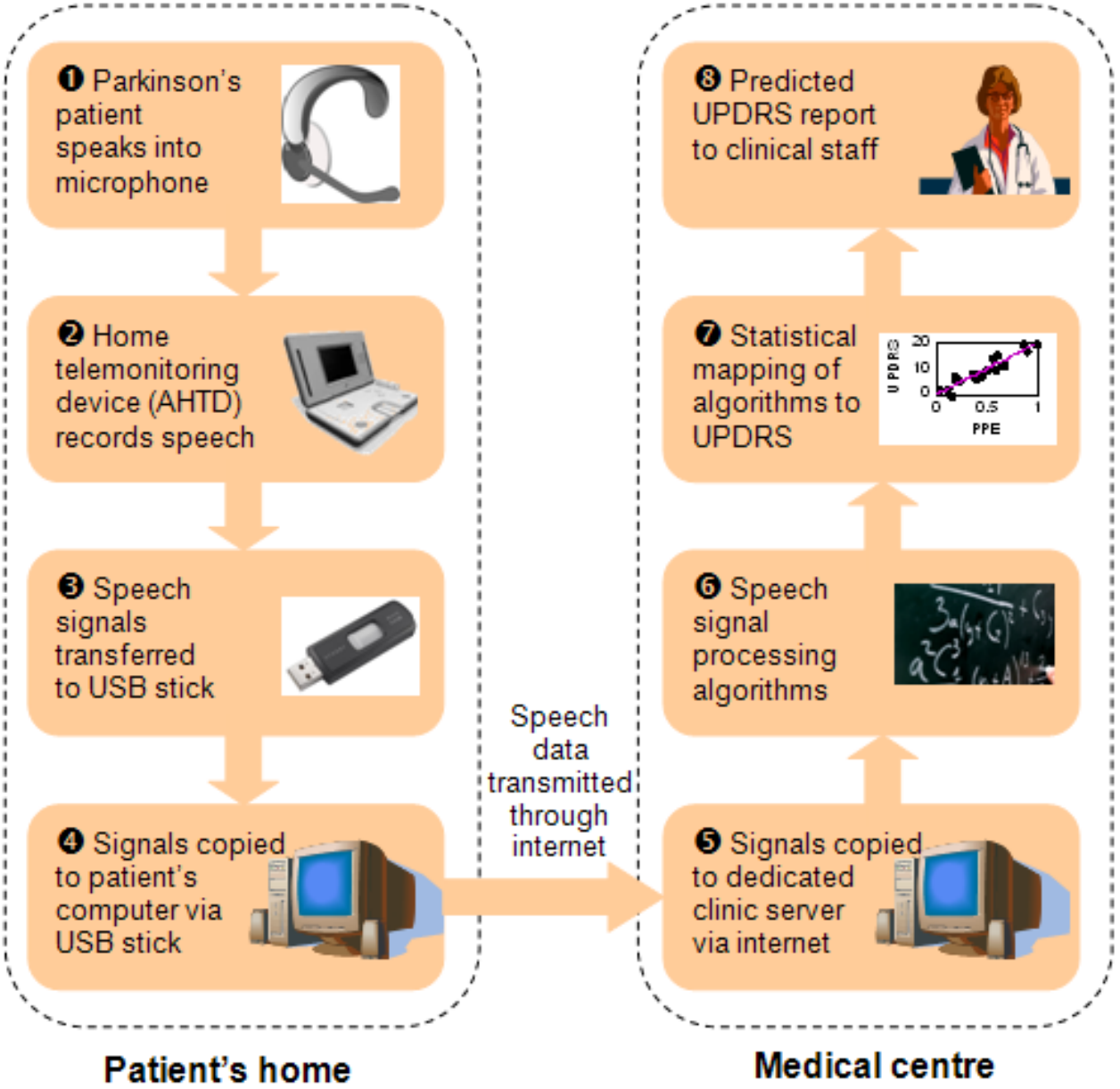


Quiz

- When constructing predictive models for identifying and measuring the severity of a medical disorder, select the correct statement.
  a) Regression for severity
  b) Classification for disorder detection
  c) All of the above are true
  d) None of the above are true

- **Slido.com #66086**

PD Telemonitoring



Telemonitoring

- Assessment of the practical value of existing traditional and nonstandard measures for discriminating healthy people from people with Parkinson's disease (PD) by detecting dysphonia.
- We found that nonstandard methods in combination with traditional harmonics-to-noise ratios are best able to separate healthy from PD subjects.

Little, McSharry, Hunter, Spielman, Ramig (2009). IEEE Trans Biomed Eng 56(4): 1019-1022.

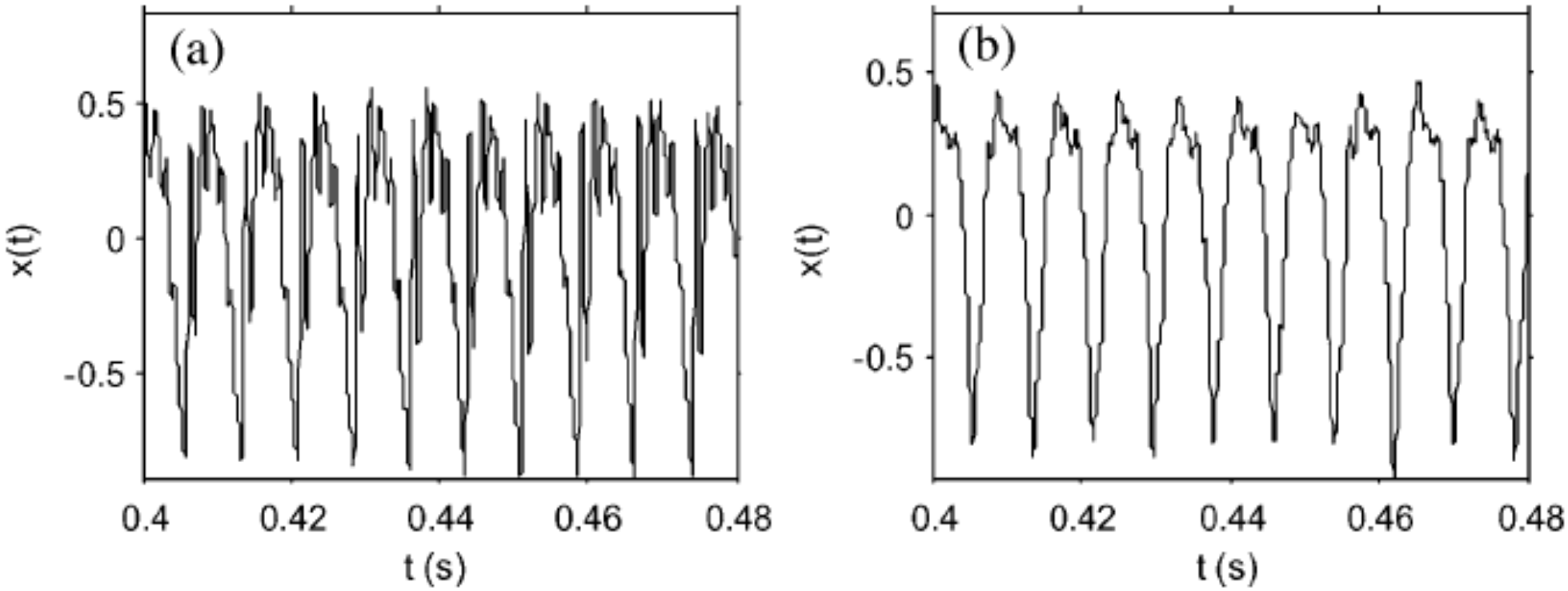
UPDRS

- The Unified Parkinson's disease rating scale (UPDRS) reflects the presence and severity of symptoms (but does not quantify their underlying causes).
- For untreated patients, it spans the range 0–176, with 0 representing healthy state and 176 representing total disabilities, and consists of three sections: 1) mentation behavior, and mood; 2) activities of daily living; and 3) motor.

Motor UPDRS

- Motor UPDRS ranges from 0 to 108, with 0 denoting symptom free and 108 denoting severe motor impairment, and encompasses tasks such as speech, facial expression, tremor, and rigidity.
- Speech has two explicit headings within Motor UPDRS and ranges between 0 and 8 with 8 being unintelligible.

Speech Signals



Two selected examples of speech signals. (a) Healthy. (b) Subject with PD.

Intuition behind DFA

- For many healthy subjects, the energy in the airflow of the lungs is well imparted to the movement of the vocal folds to generate clear sustained phonations.
- Thus, the speech signal will be smoother, and this is shown in the smaller DFA scaling exponent.
- However, many with PD are unable to maintain stable vocal fold vibration, and much more of the airflow energy will be transferred to turbulent acoustic noise generation mechanisms.
- Hence, the speech signal will be rougher, and this can be seen in an increase in the DFA scaling exponent.

Data and performance

- Collected sustained phonations from 31 people, 23 with PD.
- Selected ten highly uncorrelated measures, and an exhaustive search of all possible combinations of these measures.
- Combination of four measures leads to overall correct classification performance of 91.4%, using a kernel support vector machine.

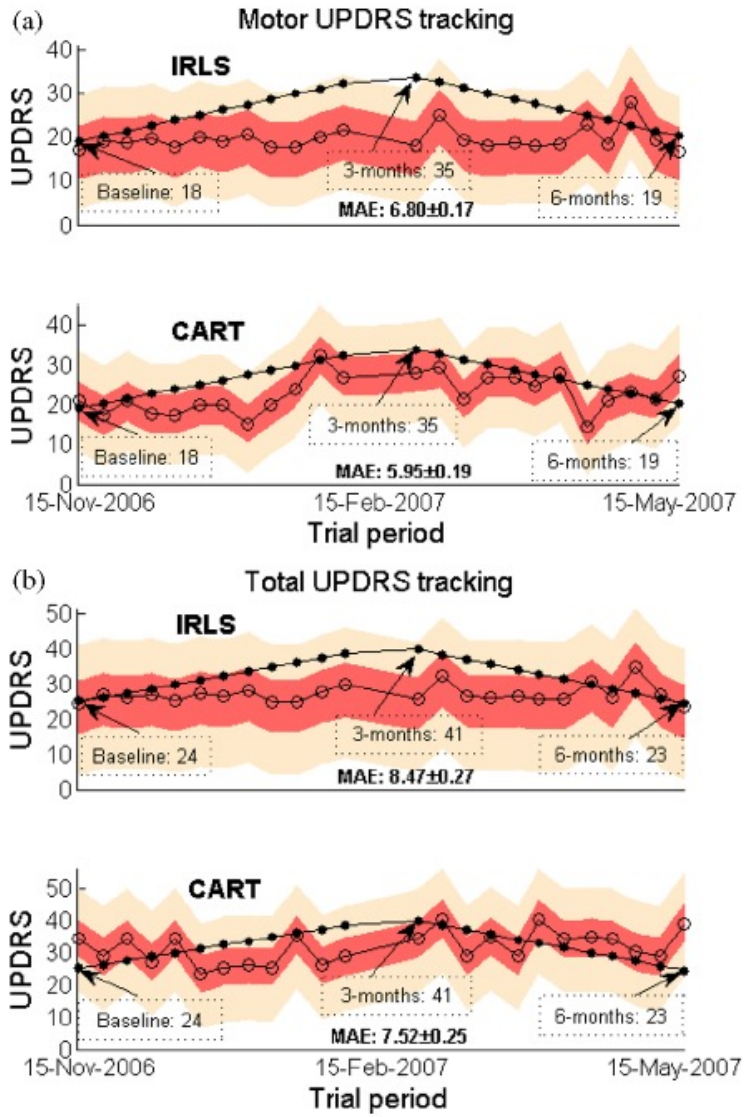
Little, McSharry, Hunter, Spielman, Ramig (2009). IEEE Trans Biomed Eng 56(4): 1019-1022.

Telemedicine monitoring of Parkinson's disease

- Intel Case Studentship investigating novel mapping of voice signals to Unified Parkinson's Disease Rating Scale (UPDRS).
- Largest database of PD with around 6000 recordings from 42 PD patients, recruited to a six-month, multicenter trial.
- Accurate telemonitoring of PD progression (2.5 % error).
- Objective, machine learning implementations.
- Non-invasive, remote monitoring, less time-consuming.
- Technology enables large scale clinical trials and facilitates design and test of novel drug treatments.

Tsanas, Little, McSharry, Ramig (2010). IEEE Trans Biomed Eng 57(4): 884-893

Tracking UPDRS



Tsanas, Little, McSharry, Ramig (2010). IEEE Trans Biomed Eng 57(4): 884-893

MAE for UPDRS

Technique	Motor-UPDRS regression MAE	Total-UPDRS regression MAE
IRLS	6.80 ± 0.17	8.47 ± 0.27
LASSO	5.95 ± 0.19	7.52 ± 0.25

Iteratively Reweighted Least Squares (IRLS) is a robust method designed to effectively reduce the influence of values distant from the main bulk of the data (outliers) by making iterative LS predictions that reweight outliers at each step.

Least Absolute Shrinkage and Selection Operator (LASSO) is a principled *shrinkage method* that acts as a powerful feature selection tool, which also offers a mathematical framework enhancing the physiological interpretability of the resulting regression coefficients. LASSO has the desirable characteristic of simultaneously minimizing the prediction error whilst producing some coefficients that are effectively zero (reducing the number of relevant input variables) by adjusting a *shrinkage* parameter.

PD diagnosis

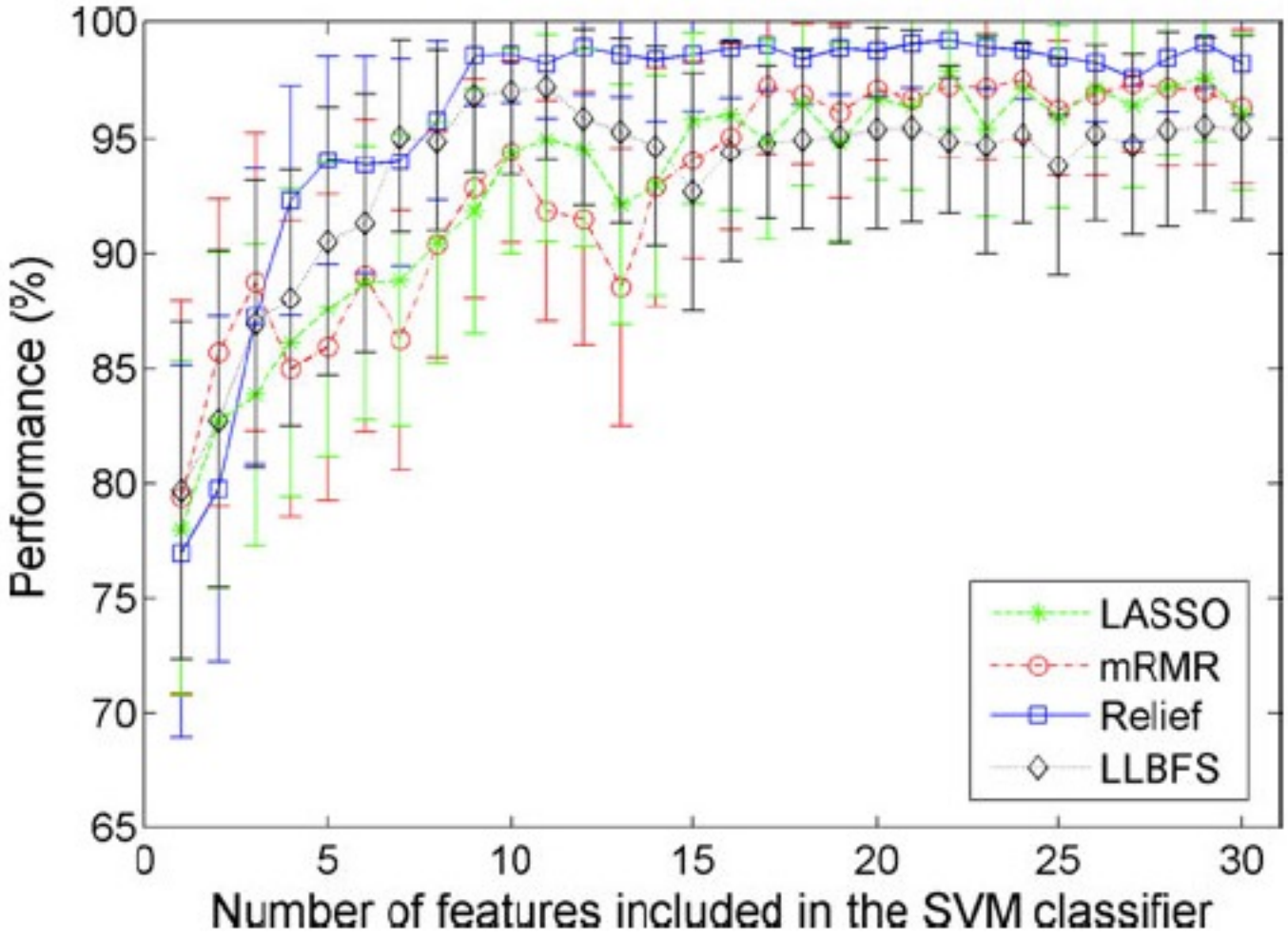
- Database containing 43 PD subjects.
- Computed 132 dysphonia measures from sustained vowels.
- Parsimonious subsets of these dysphonia measures using four feature selection algorithms.
- These feature subsets were mapped to a binary classification response.
- Two statistical classifiers were assessed: random forests and support vector machines.

Tsanas, Little, McSharry, Ramig (2012). IEEE Trans Biomed Eng 59(5): 1264-1271

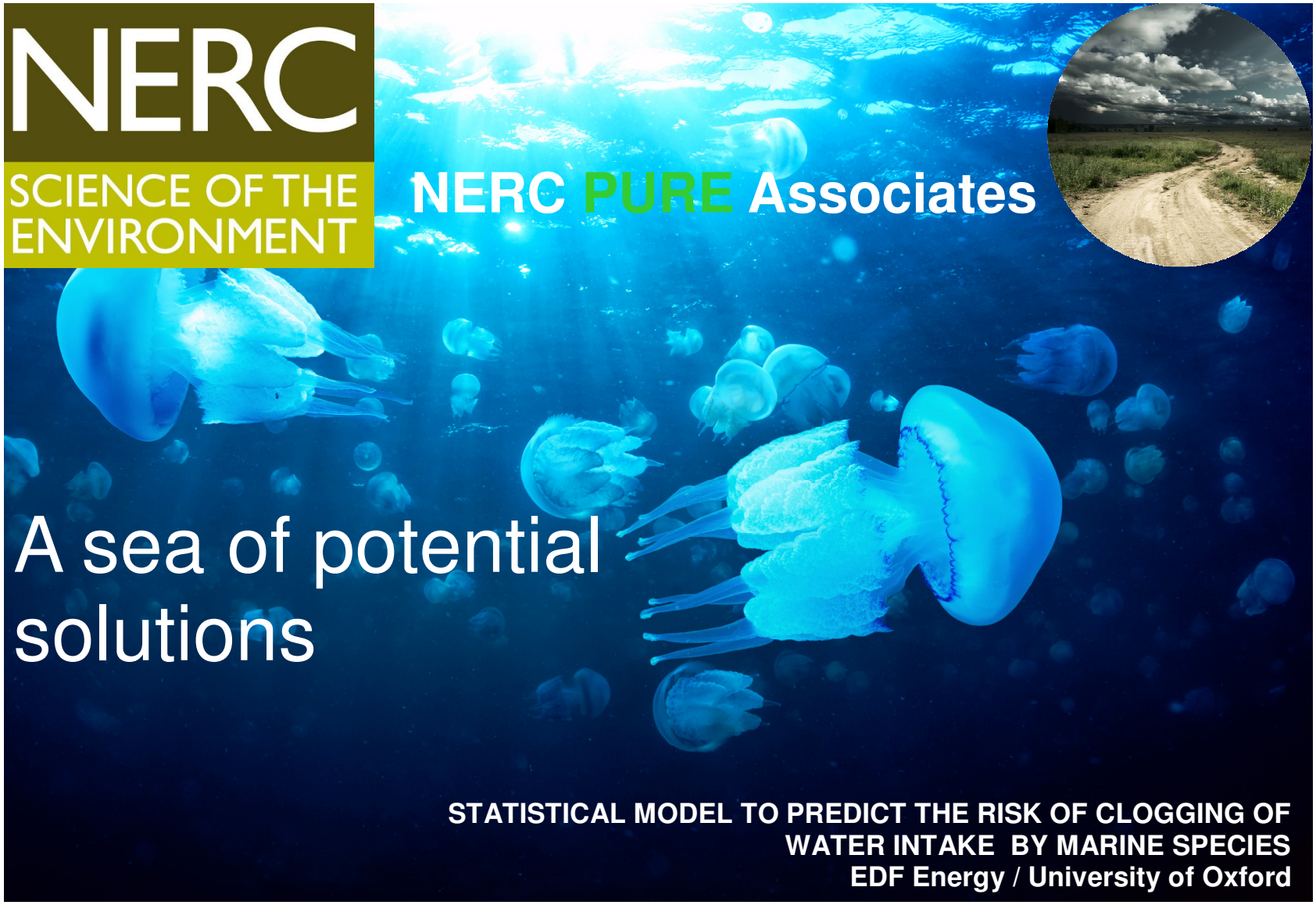
Results

- An existing database consisting of 263 samples from 43 subjects.
- New dysphonia measures can outperform state-of-the-art results.
- Almost 99% overall classification accuracy was obtained using only ten dysphonia features.
- SVM was was more accurate (97.7 ± 2.8) than RF (90.2 ± 5.9) when using all 132 features.

Performance using SVM



Nuclear reactor clogging



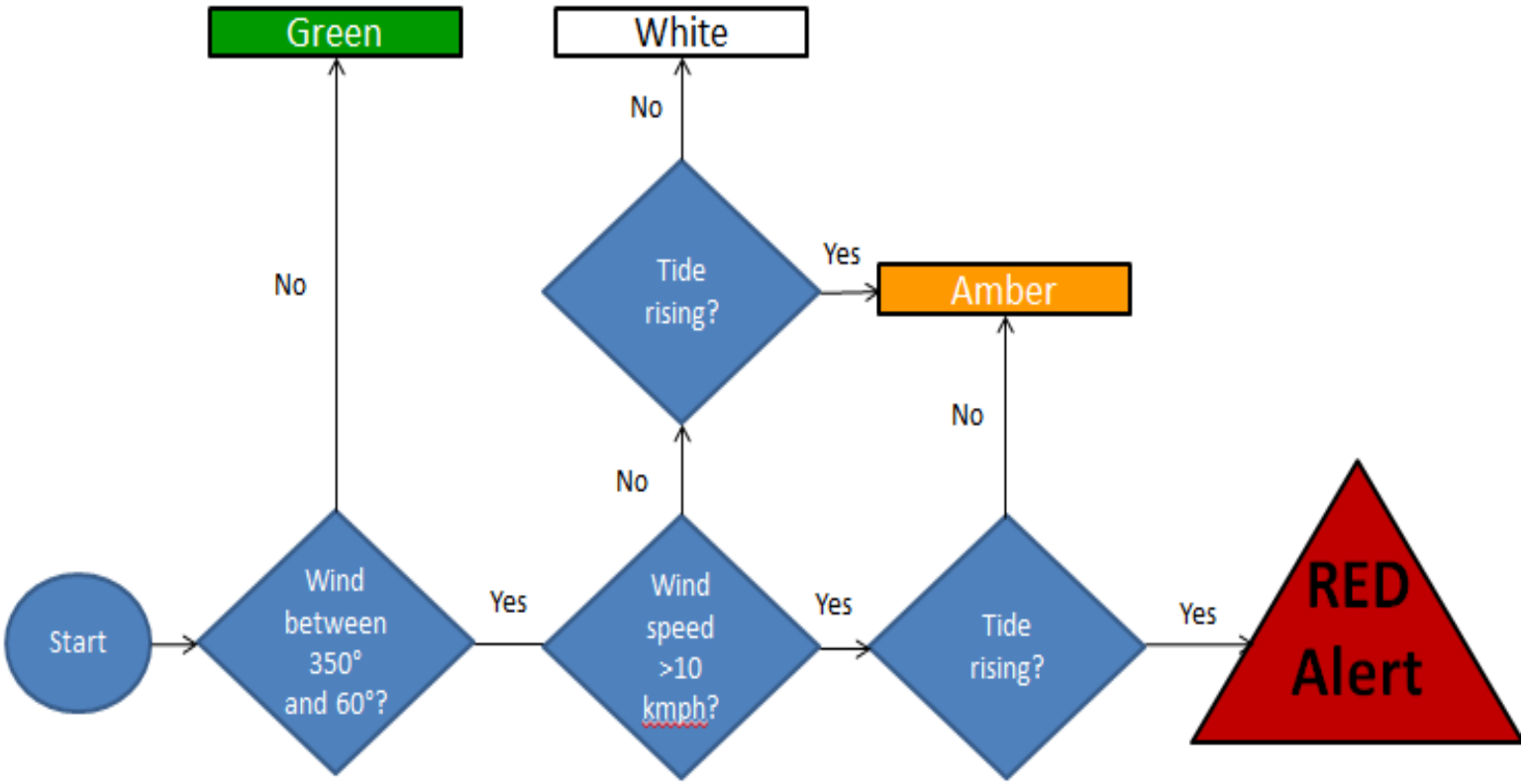
The need

- Across the past few decades in North Western Europe, in particular along the Atlantic and the North Sea coast, massive occurrences of jellyfish and seaweed have been observed.
- These marine species can block power plant cooling systems, reducing their performance, or even stopping them from pumping sea water all together.
- Some of these events have led to the halt of electricity production for up to a week, with costly consequences, for example, at Torness Nuclear Power Station, Scotland.
- These clogging incidents appear difficult to predict, mainly due the historical scarcity of such events, which makes it very difficult to pinpoint the driving factors behind them.

The challenge

- The challenge was to create an Early Warning System (EWS) for clogging and ingress events at Torness.
- The goal was to provide a model that would be simple to use and that would lay the foundations for an effective tool to be further developed by EDF.
- The project succeeded in gathering and reconstructing the necessary data for the primary analysis.
- This enabled us to identify the marine and weather factors that drive the excessive presence of marine debris in seawater, eventually causing ingresses.

Traffic light alert system

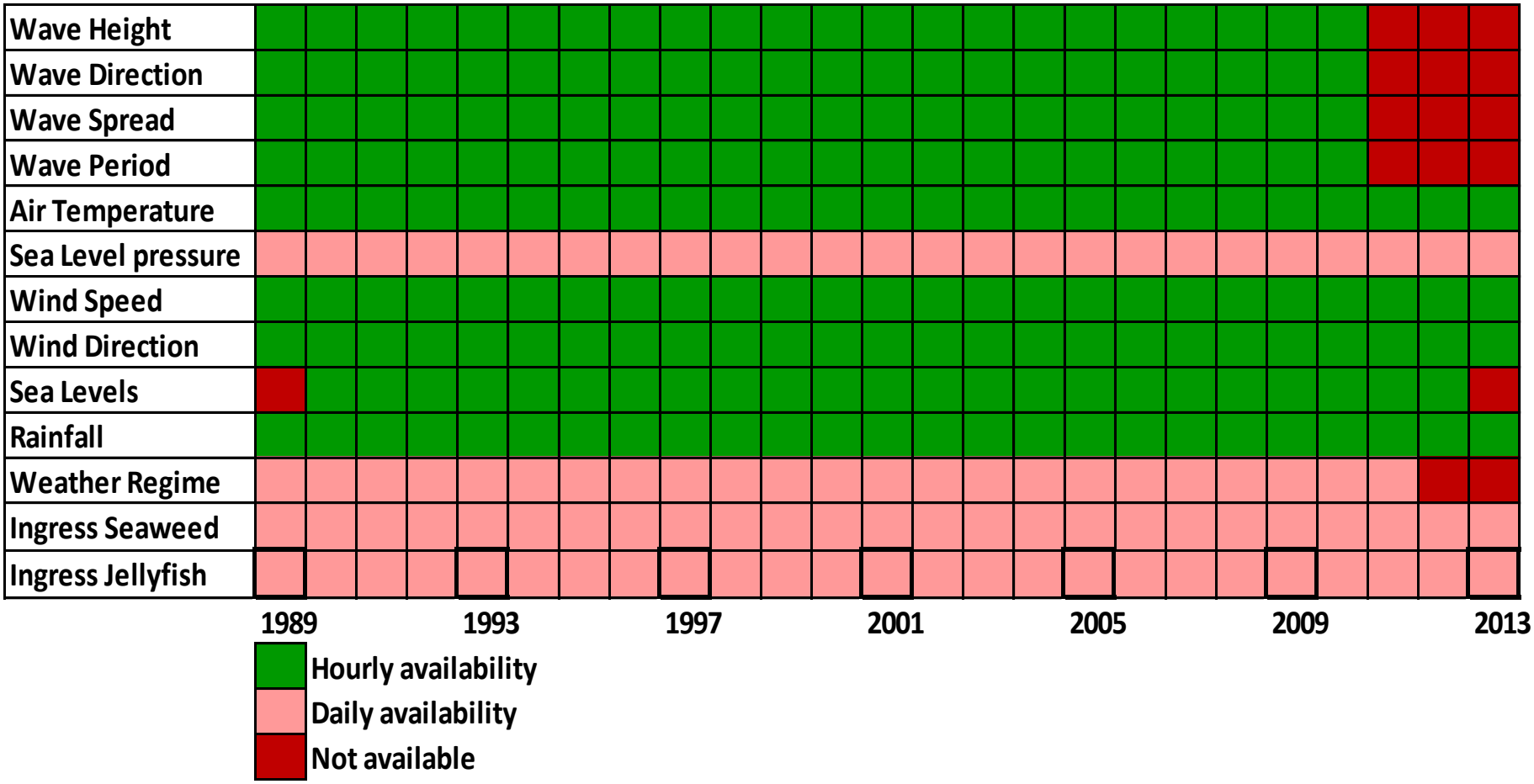


This classifier based on wind direction, wind speed and tide served as the benchmark model.

Data availability

Source	Data	Frequency	Missing Data?	Time Span
EDF	Wave Height	Hourly	No	1979 – 2010
EDF	Wave Direction	Hourly	No	1979 – 2010
EDF	Wave Spread	Hourly	No	1979 – 2010
EDF	Wave Period	Hourly	No	1979 - 2010
	Air Temperature	Hourly	No	1979 – 2012
	Sea Level pressure	Daily	No	
BADC	Wind Speed	Hourly	Yes	1979 – 2013
BADC	Wind Direction	Hourly	Yes	1979 – 2013
BADC	Sea Levels	Hourly	Yes, a lot	1993 – 2012
BADC	Rainfall	Hourly	Yes	1979 – 2013
MetOffice	Weather Regime	Daily	No	1850 – 2011
Torness	Ingress Seaweed	Daily	-	
Torness	Ingress Jellyfish	Daily	-	

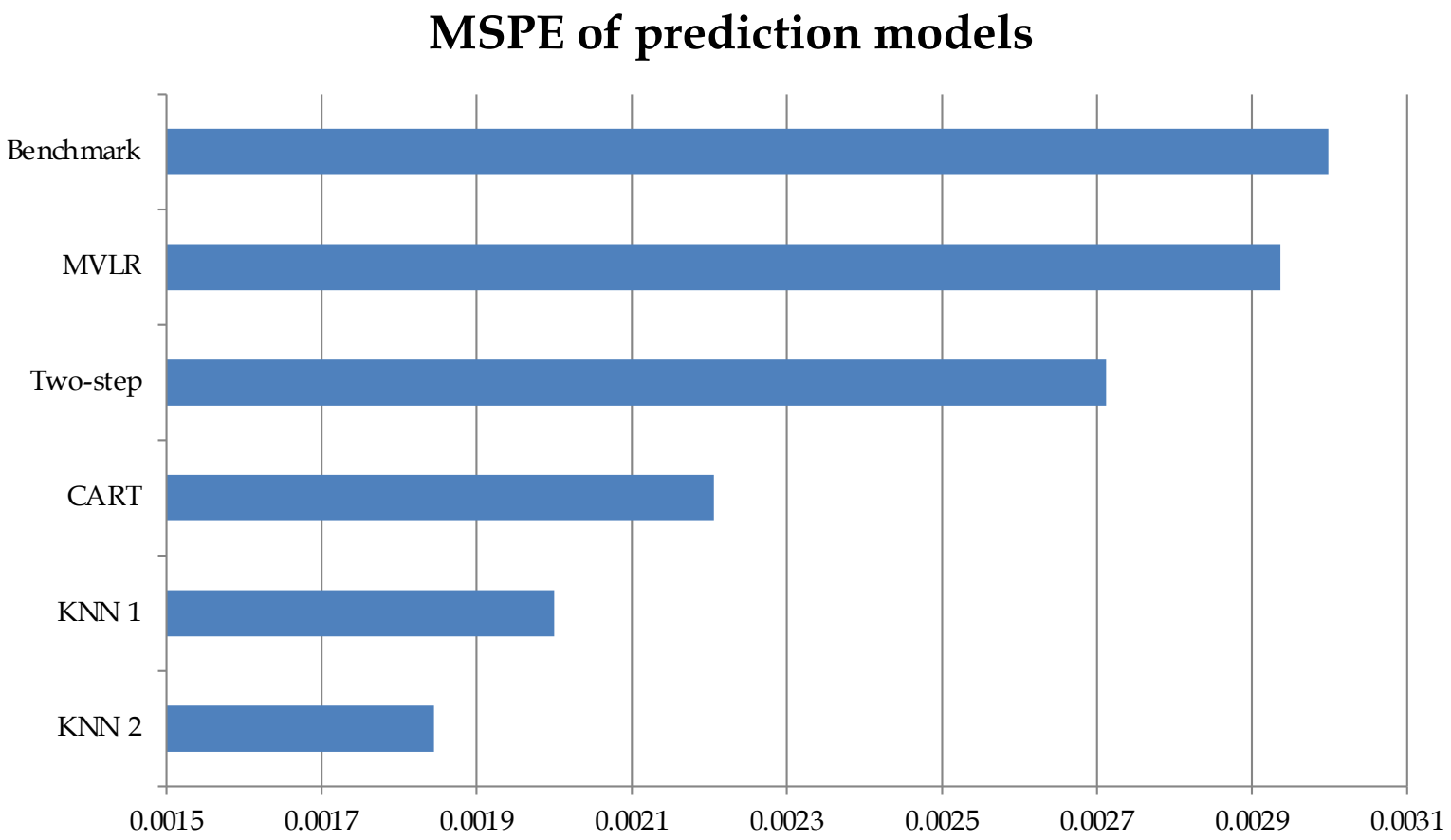
Temporal representation of data



Key results

- Weather variables, marine variables and weather regime patterns are all responsible for ingresses.
- Identification of lead-lag relationships was important for maximizing correlation.
- The successful results obtained by the CART model and K-Nearest-Neighbors (KNN) model suggest that a machine learning approach provides the best performance.
- CART and KNN lowered the Mean Squared Prediction Error (MSPE) by 26% and 40% respectively in comparison to the benchmark.

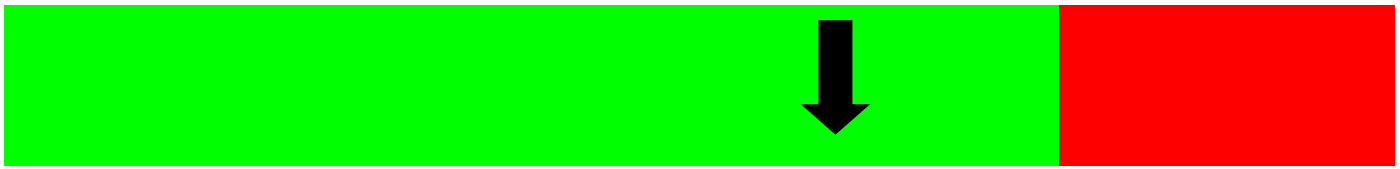
Model comparison



Risk communication

Continuous scale, with no pre-defined threshold:

Quantised scale with pre-defined threshold:



Low risk

High risk

Poll: Risk communication

- When comunicating risk to society, how many risk levels would you recommend?
  a) Two
  b) Three
  c) Four
  d) Five

- **Slido.com #66086**

Probabilistic early warning

	Green	Yellow	Amber	Red	
Warning	None	Advisory	Advisory	Early	Flash
Risk	Very low <20%	Low ≥20% <40%	Moderate ≥40% <60%	High ≥60% <80%	Very high ≥80%
Headline	No extreme event expected	Low risk of extreme event	Moderate risk of extreme event	High risk of extreme event	Extreme event is imminent or occurring
Impact		Low risk of major damage to infrastructure	Moderate risk of major damage to infrastructure	High risk of major damage to infrastructure. Casualties are possible	Major damage to infrastructure is likely. Casualties are possible
Advice		Ensure you access the latest risk forecast	Remain vigilant and ensure you access the latest risk forecast	Remain extra vigilant and access the latest risk forecast. Be aware of risks that might be unavoidable. Follow any advice given by authorities	Remain extra vigilant and access the latest risk forecast. Follow orders and any advice given by authorities under all circumstances and be prepared for extraordinary measures

Following UK Met Office severe weather warnings.

Data Analytics

WEEK 6B

Course outline

Week	Lecture A	Lecture B
1	Data Analytics	Weather forecasting
2	Renewable energy	Wind energy
3	Solar energy	Demand forecasting
4	Risk	Extreme events
5	Health	Biomedicine
6	Early warning systems	Economic forecasting

Today's Lecture

No.	Activity	Description	Time
1	Challenge	Economic activity	10
2	Discussion	Expansion and recession	10
3	Case study	US GNP forecasting	10
4	Analysis	Regime switching	20
5	Demo	Parsimonious models	20
6	Q&A	Questions and feedback	10

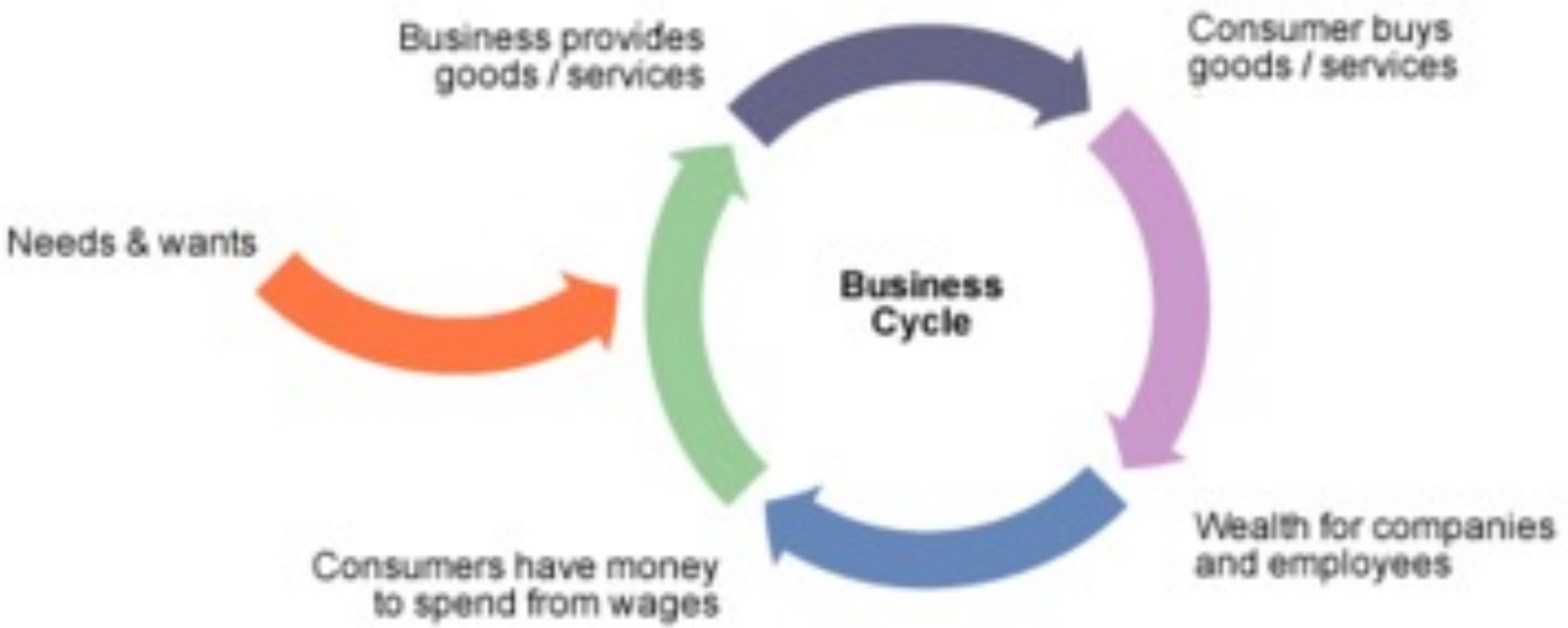
Economic activity

- Economic activity refers to actions that involve the production, distribution and consumption of goods and services at all levels within a society.
- Gross domestic product or GDP is one way of assessing economic activity.
- Current economic activity and forecasts for its future level can significantly impact business activity and profits, as well as inflation and interest rates.

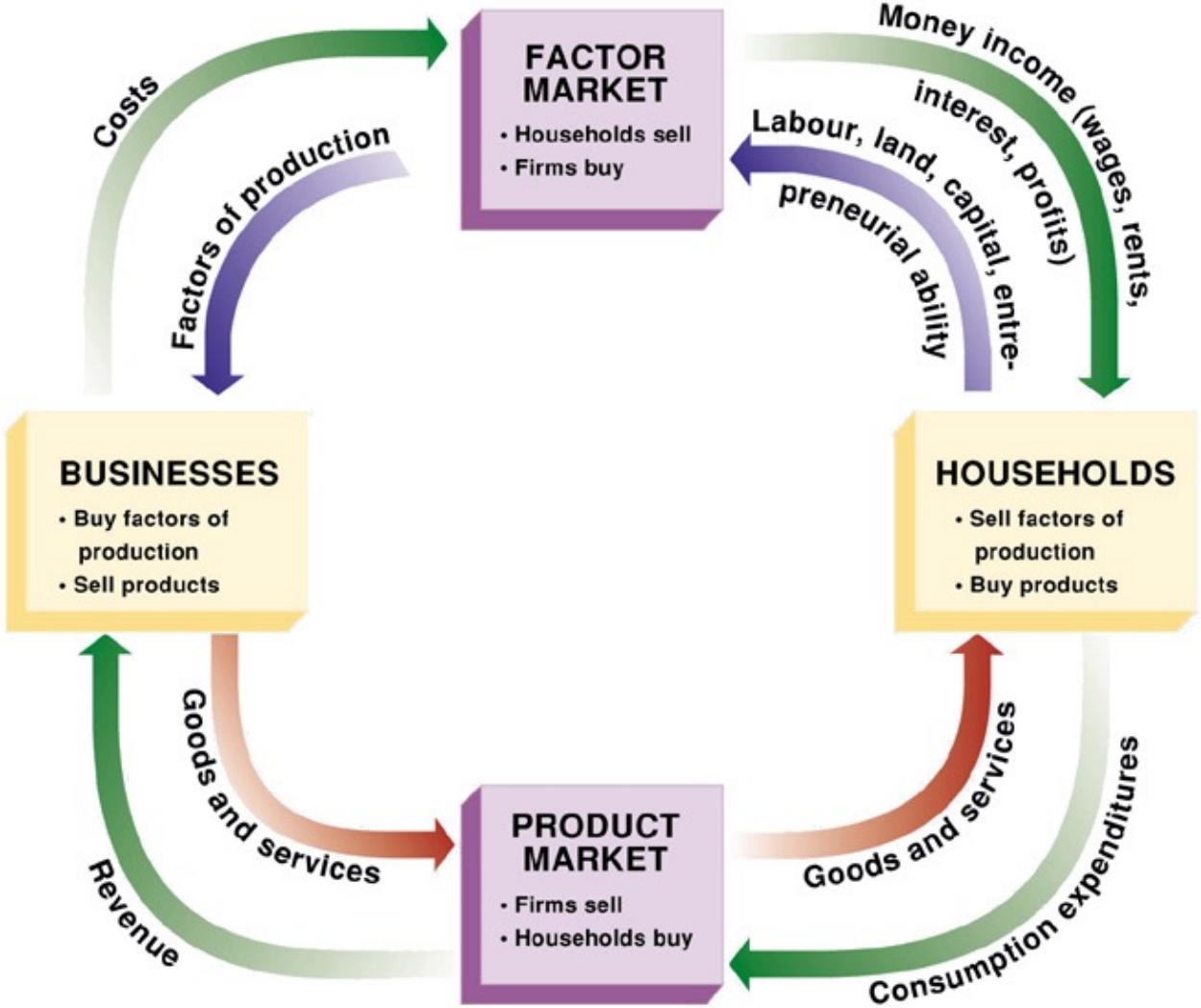
Economic activity levels

	Primary	Secondary	Tertiary	Quaternary
Natural Resources	Use natural resources directly	Process natural resources	Do not directly gather or process raw materials	Do not need to be located near resources or a market
Examples	Farming Mining Forestry Fishing	Processing flour from wheat Making lumber from trees Producing electrical power	Doctors Salespeople Firefighters Truck drivers	Government Education ICT Research

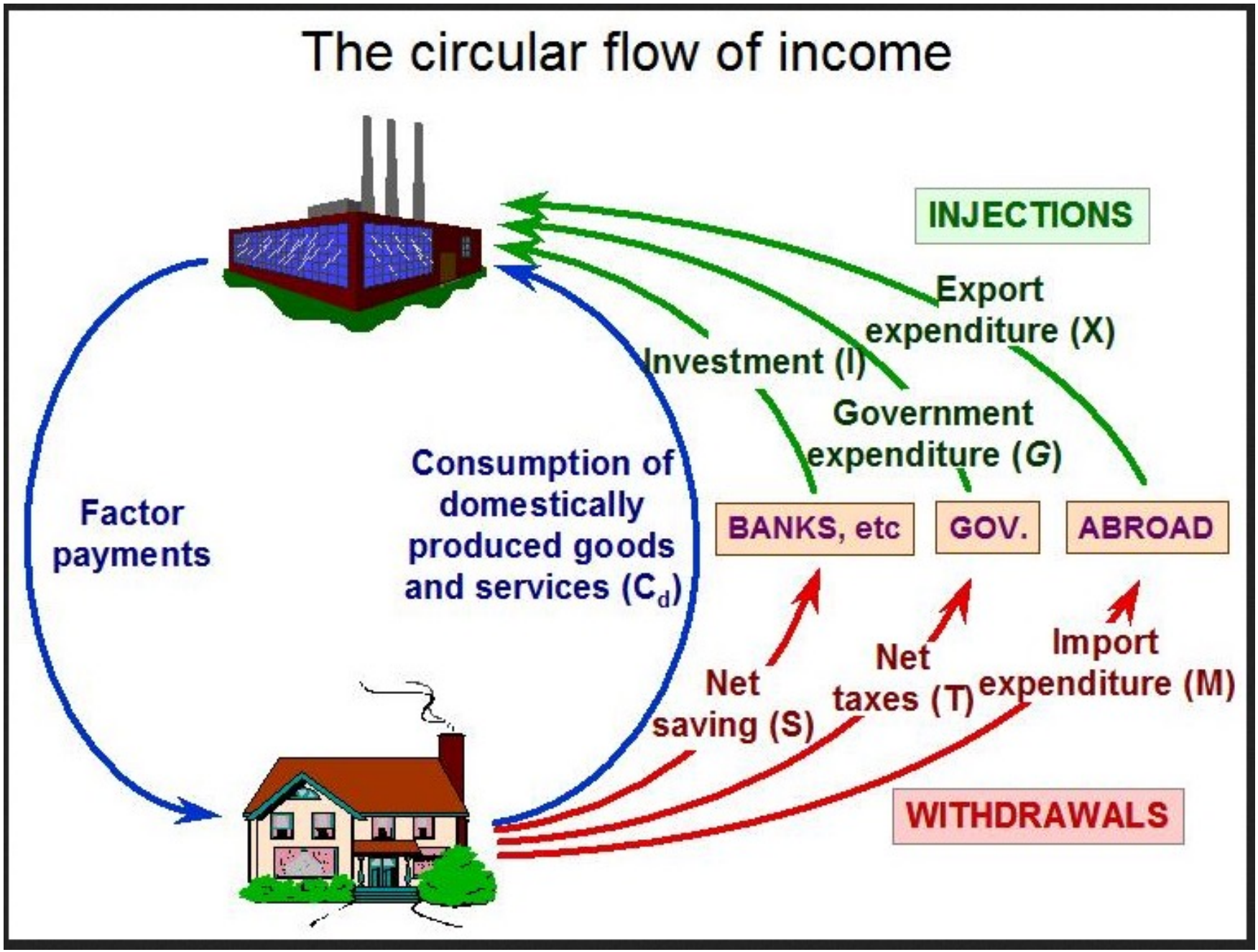
Business cycle



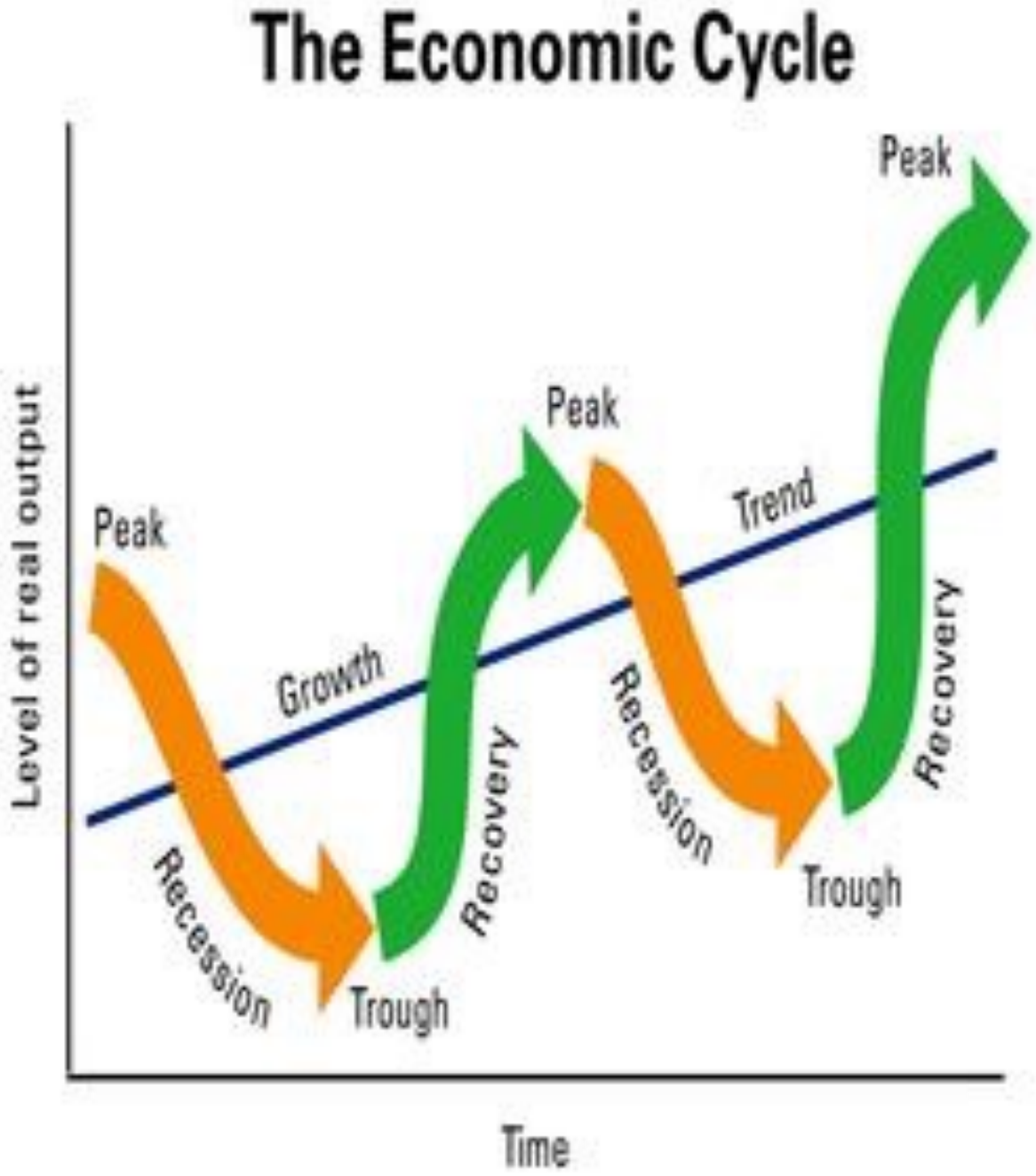
Circular Flow Diagram



Source: wiki.ubc.ca

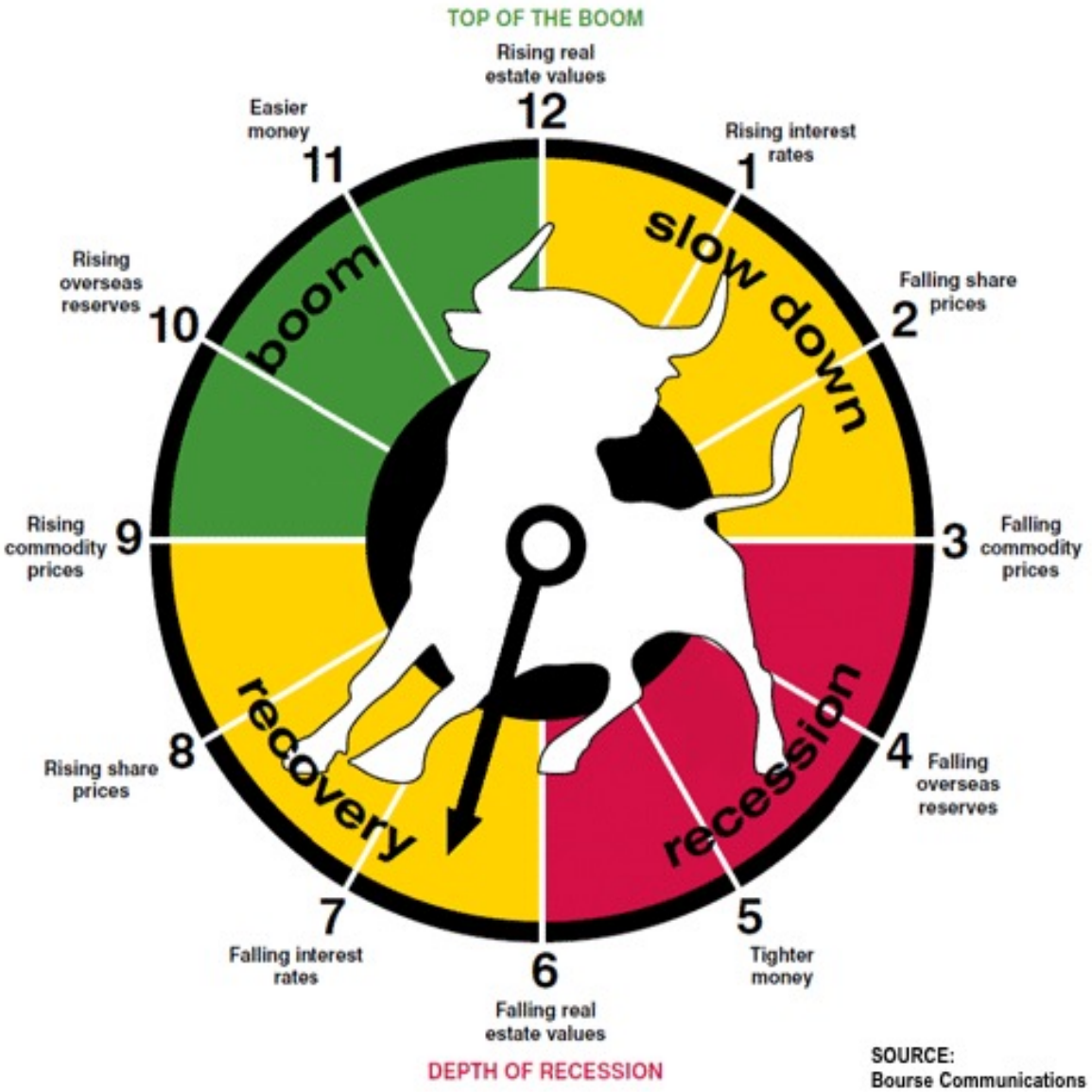


Source: Edward Lambert

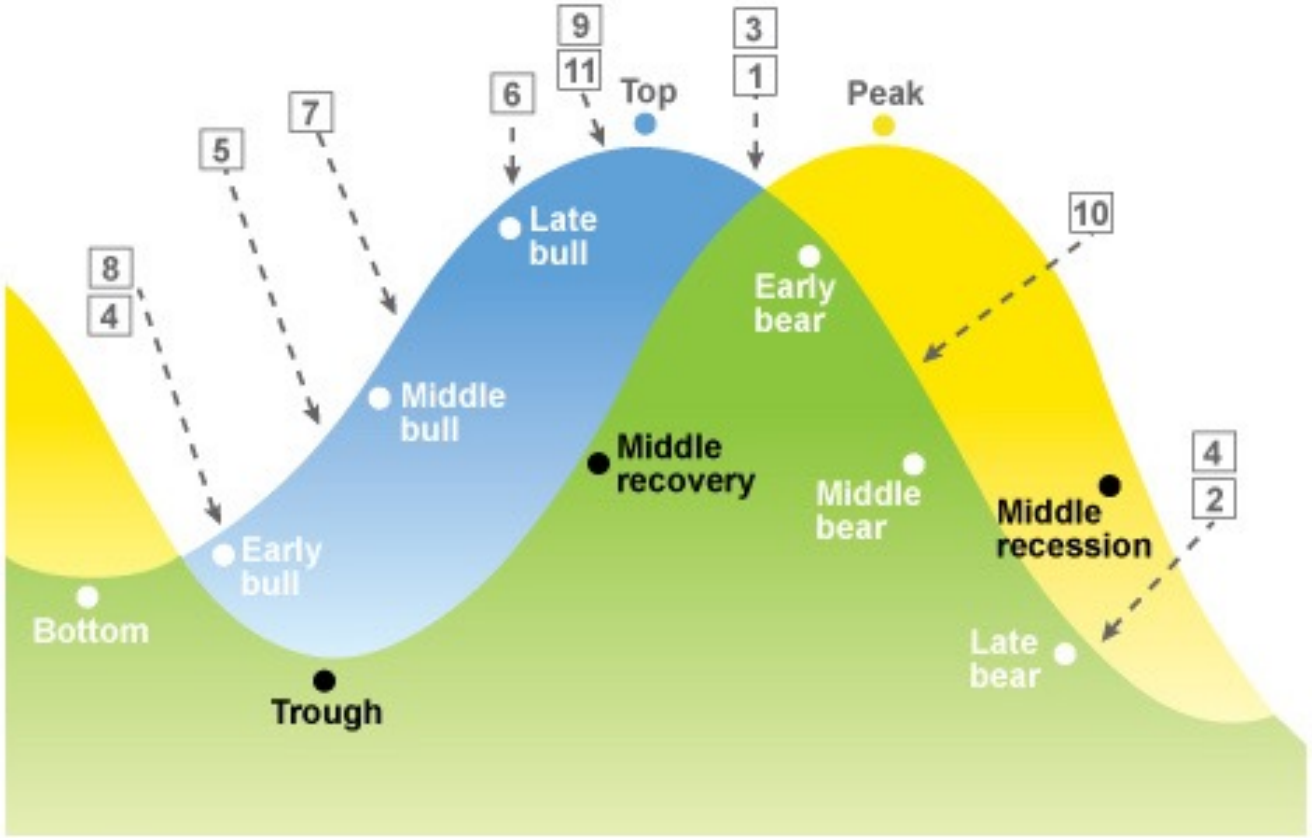

The Economic Cycle
Level of real output
Time
Peak
Peak
Peak
Trend
Growth
Recession
Recovery
Recession
Recovery
Trough
Trough

Economic Cycles

The cycle


TOP OF THE BOOM
Rising real estate values
12
Rising interest rates
1
Falling share prices
2
Falling commodity prices
3
Falling overseas reserves
4
Tighter money
5
Falling real estate values
6
Falling interest rates
7
Rising share prices
8
Rising commodity prices
9
Rising overseas reserves
10
Easier money
11
boom
slow down
recovery
recession
DEPTH OF RECESSION
SOURCE:
Bourse Communications

Business Cycle and Relative Stock Performance

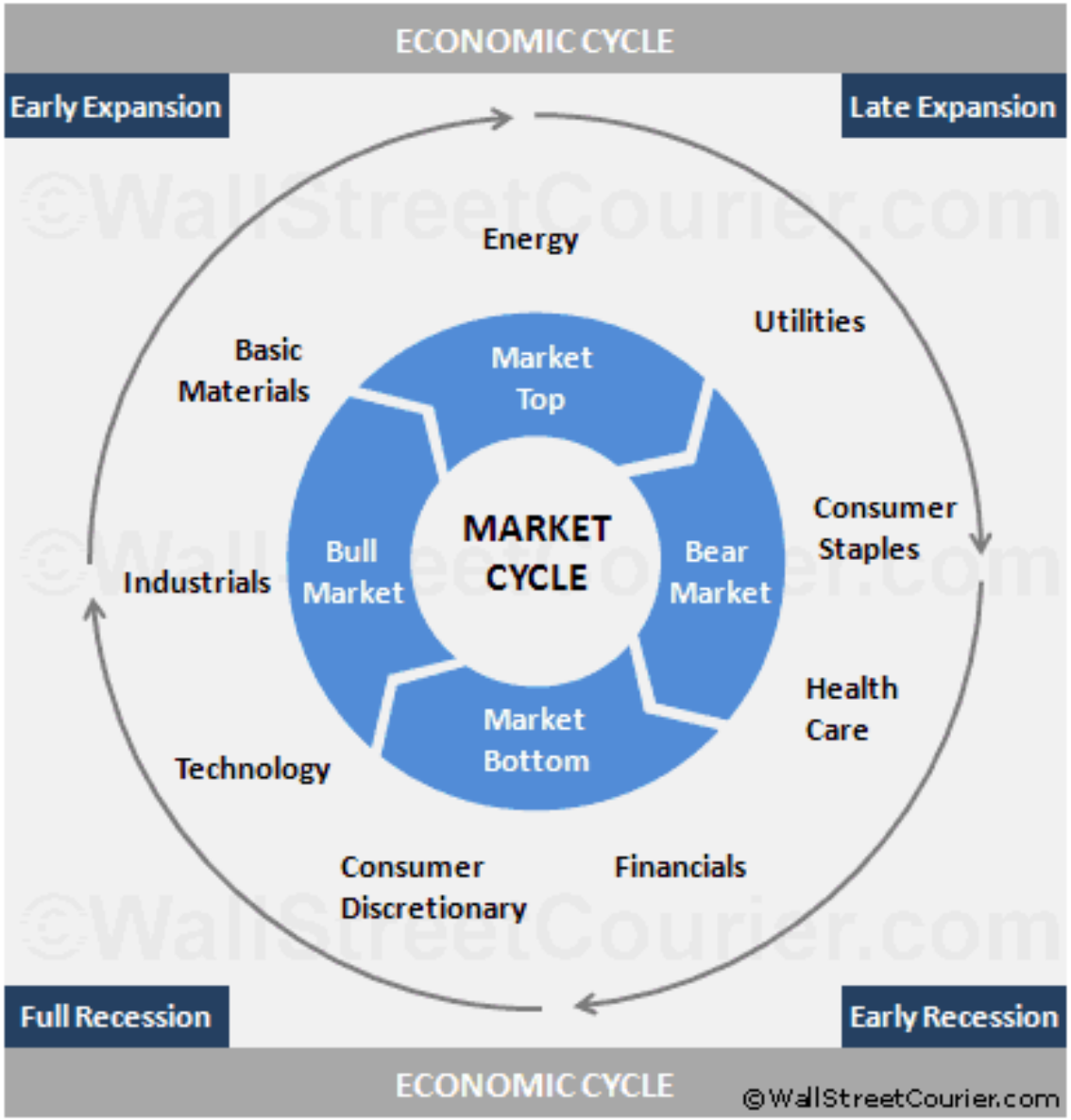


Key ● Economic Cycle ● Stock Market Cycle

1 Consumer Non-Cyclicals
2 Consumer Cyclicals (durable & non)
3 Health Care
4 Financials
5 Technology
6 Basic Industry
7 Capital Goods
8 Transportation
9 Energy
10 Utilities
11 Precious metal

Source: Stovall, Sam. The S&P Guide to Sector Investing. New York: McGraw-Hill, 1995.

Economic clock and sectors



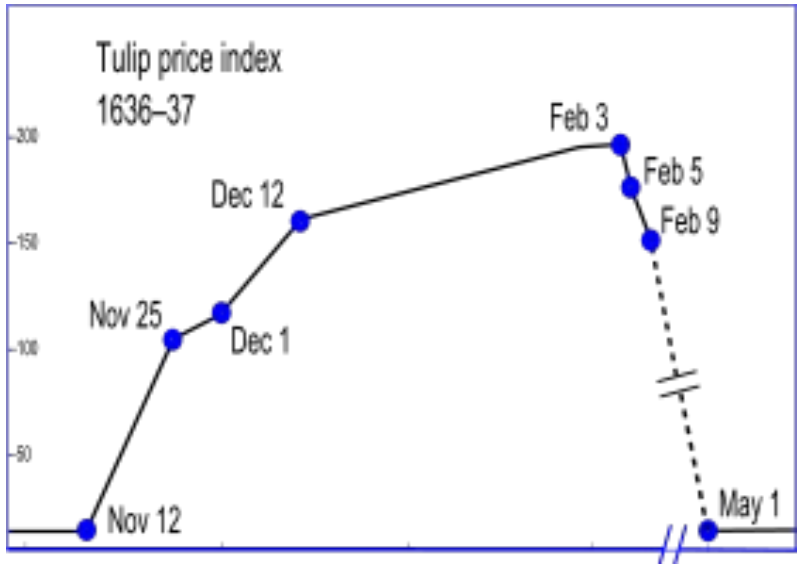
Quiz

- Which of the following is viewed as the first speculative financial bubble?
  a) Dot-com bubble
  b) Subprime crisis
  c) South Sea bubble
  d) Tulip mania

- **Slido.com #81967**

Tulip prices

- In Europe, formal futures markets appeared in the Dutch Republic during the 17th century.
- One of the most notable was the tulip market.
- At the peak of tulip mania, in February 1637, some single tulip bulbs sold for more than 10 times the annual income of a skilled artisan.



Major economic shocks

- Tulip mania (1637)
- South Sea Bubble (1720)
- Wall Street Crash of 1929
- Great Depression (1929–1939)
- Black Monday (1987)
- Dot-com bubble (2000s)
- Subprime mortgage crisis (2007-2009)
- Eurozone crisis (2009)
- Chinese stock market crash (2015)
- COVID pandemic (2020)

Bond ratings

Moody's Rating	Average default rate within one year of rating (1970-2001)	Definition	Notes
Aaa	0.00%	Highest rating available	Investment grade bonds
Aa	0.02%	Very high quality	
A	0.01%	High quality	
Baa	0.15%	Minimum investment grade	
Ba	1.21%	Low grade	Junk bonds
B	6.53%	Very speculative	
Caa	24.73%	Substantial risk	
Ca		Very poor quality	
C		Imminent default or in default	

Source: www.blaha.net

Credit crisis

- Exposures to credit risk via subprime mortgages and credit default swaps
- Lehman Brothers Holdings, Inc. was rated A2 by Moody's and A by Standard & Poor's until it filed for bankruptcy on September 15, 2008
- American International Group (AIG) was also rated A2 by Moody's and A- by S&P when it received an $85 billion loan from the federal government in exchange for an 80% stake on September 16, 2008 to stay afloat

Poll on GDP

- Gross Domestic Product (GDP) measures the amount of economic activity per person is a given country.

a) True
b) False

- **Slido.com #66086**

Gross Domestic Product

- Gross Domestic Product (GDP) is the broadest quantitative measure of a nation's total economic activity.
- More specifically, GDP represents the monetary value of all goods and services produced within a nation's geographic borders over a specified period of time.
- GDP is typically reported on a quarterly time frame and may be given by economic sector.

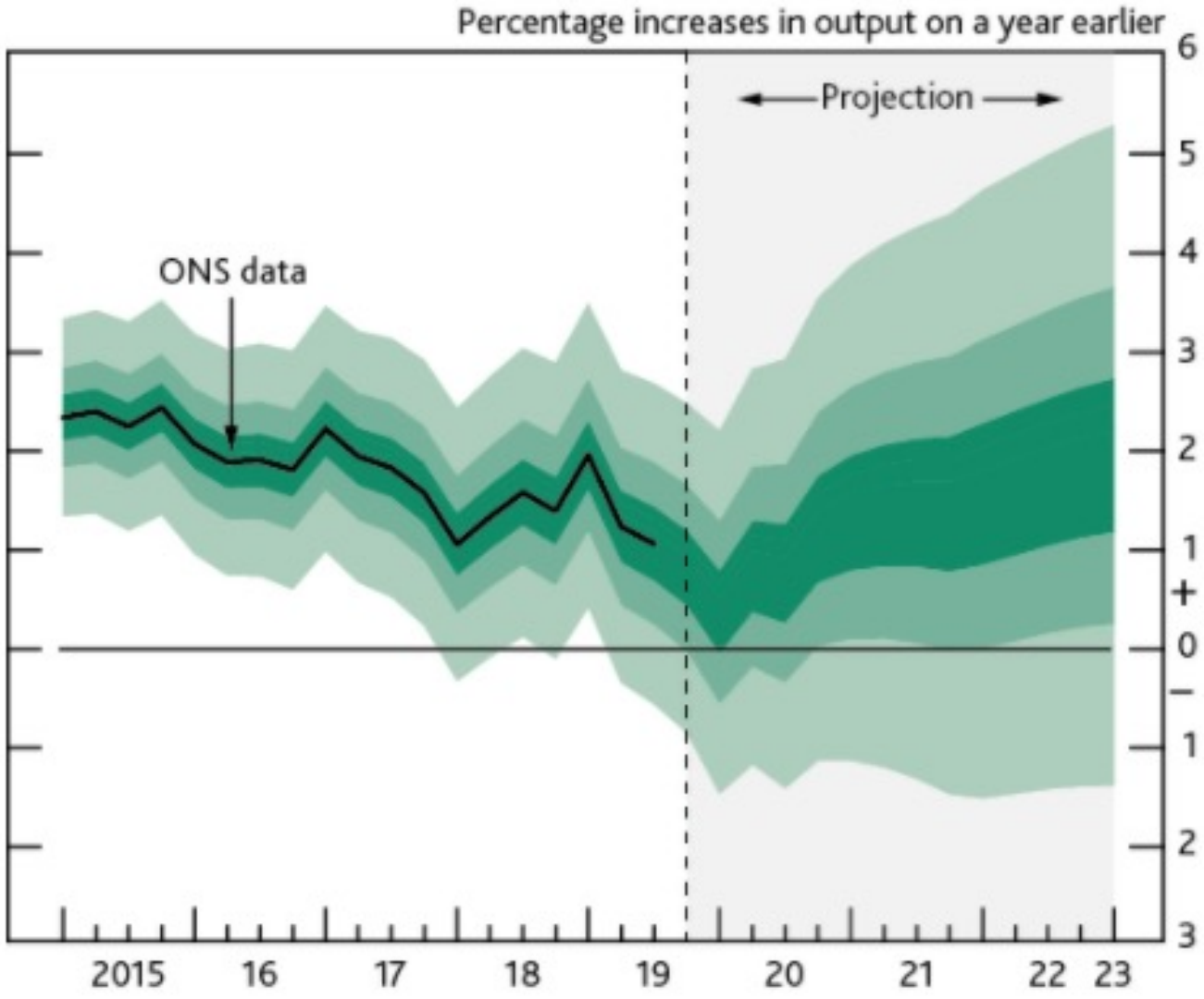
GDP versus GNP

- Gross Domestic Product (GDP) and Gross National Product (GNP) are closely related measures.
- GDP measures the total output of the economy in a period i.e. the value of work done by employees, companies and self-employed persons.
- This work generates incomes but not all of the incomes earned in the economy remain the property of residents (and residents may earn some income abroad).
- The total income remaining with the residents is the GNP and it differs from GDP by the net amount of incomes sent to or received from abroad.

Bank of England

- Approaches for describing the evolution of uncertainty have also been adopted by economic institutions such as the Bank of England.
- Uncertainty in GDP and inflation is represented via Monetary Policy Committee's fan charts
- Uncertainty arises from the range of different opinions provided by human experts.
- Charts represent the best collective judgment of the Committee, conditional on a particular path for interest rates.

Bank of England GDP fan chart



GNP model structures

Benchmarks	• Persistence • Unconditional Average
Linear	• AR(4)
Nonlinear & parametric	• SETAR • MS-AR
Nonlinear & non-parametric	• F-NN • Kernel Regression, WRAP

Poll

- Which model do you think will be best for forecasting GNP?
  a) Benchmarks
  b) Linear
  c) Nonlinear & parametric
  d) Nonlinear & nonparametric

- **Slido.com #66086**

Regime switching

- Intuition about the economy suggests the existence of (at least) two regimes
- Regime 1: expansion (economic growth)
- Regime 2: contraction (recession)
- Given that the economy is in a particular regime, use a specific AR model
- A variety of model structures allow switching between regimes

SETAR(2,p,p) model

- The simplest SETAR models provides evolution of a process governed by switching between two distinct linear auto-regressions:

$$y_t = (1 - I(y_{t-d} > r))(\alpha_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \sigma_1 \varepsilon_t)$$
$$+ I(y_{t-d} > r)(\beta_0 + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \sigma_2 \varepsilon_t)$$

- $I(y_{t-d}>r) = 1$ if $y_{t-d}>r$ and zero otherwise and $\varepsilon_t \sim N(0,1)$
- Two regimes: AR(p) with coefficients $(\alpha_0,\alpha_1,\ldots,\alpha_p,\sigma_1)$ and $(\beta_0,\beta_1,\ldots,\beta_p,\sigma_2)$ and regime parameters r and d

SETAR estimation

- Split the sample into two regimes and perform OLS is each separately
- d takes on integer values 0,1,2,…
- r is allowed to take on each value y_{t-d}
- If p is known select pair (r,d) that minimises the overall sum of the squares in each regime
- If p is unknown, use Akaike (AIC) and minimise:

$$N_L \ln \hat{\sigma}_L^2 + N_U \ln \hat{\sigma}_U^2 + 2(p+1) + 2(p+1)$$

MSMean(2)-AR(4) for GNP

- Hamilton (1989) applied MS-AR to quarterly percentage changes in real US GNP:

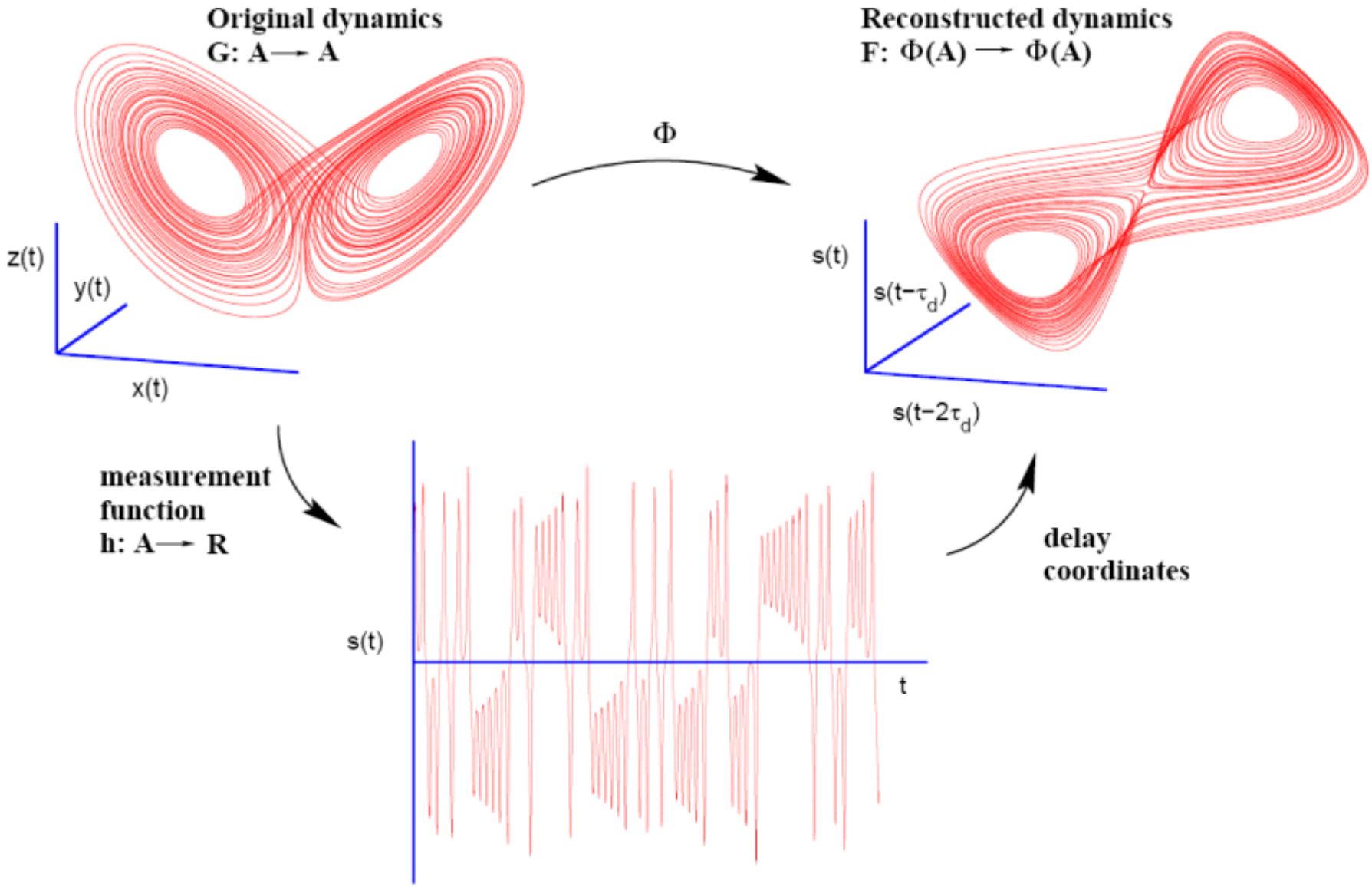
$$\Delta y_t - \mu(s_t) = \sum_{i=1}^{p} \phi_i \{\Delta y_{t-i} - \mu(s_{t-i})\} + \varepsilon_t, \quad \varepsilon_t \sim IID(0, \sigma^2)$$

- where the conditional mean $\mu(s_t)$ takes on values $\mu_1>0$ if $s_t=1$ ("boom") and $\mu_2<0$ if $s_t=2$ ("recession")
- The variance is assumed to be the same in each regime
- The parameters of the AR process depend upon the regime which is governed by the unobservable state s_t
- This model is ideal for modelling a process with regime shifts

Nonlinear dynamical systems

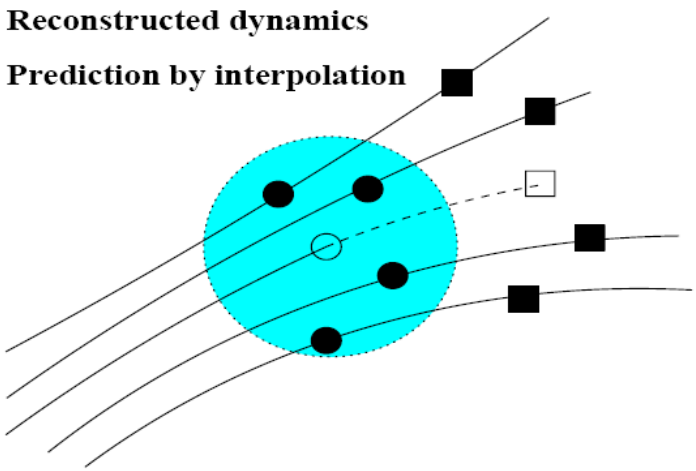
- Motivated by dynamical systems theory
- Used for modelling low-dimensional nonlinear systems (potentially chaotic systems)
- Define state space using m-dimensional delay embedding:
- $\mathbf{x}(t) = [y(t), y(t-\tau), \ldots, y(t-(m-1)\tau)]$
- Dynamics in reconstructed state space $\mathbf{x}(t)$ can provide a faithful representation of original state space

State space reconstruction



f-NN

- Fraction of nearest neighbours
- Neighbourhood formed using a ball which captures a fraction f of the available observations
- Radius at each local neighborhood adapts with the density of data in the state-space



Kernel Regression

- An extension of nearest neighbour methods
- Relies on the assumption that the more similar the past state is to the current state, the more likely it becomes for their trajectories to match
- It assigns relatively higher weights to the closest state vectors

WRAP

- Weighted random analogue prediction
- Hybrid of f-NN and kernel regression
- Requires the estimation of two parameters:
- f quantifies the size of the neighbourhood
- a bandwidth parameter β, that controls the width of the kernel employed for weighting state vectors within the neighbourhood

Poll: Model parsimony

- Which of the models has the least number of parameters?
  a) Linear
  b) Nonlinear & parametric
  c) Nonlinear & nonparametric

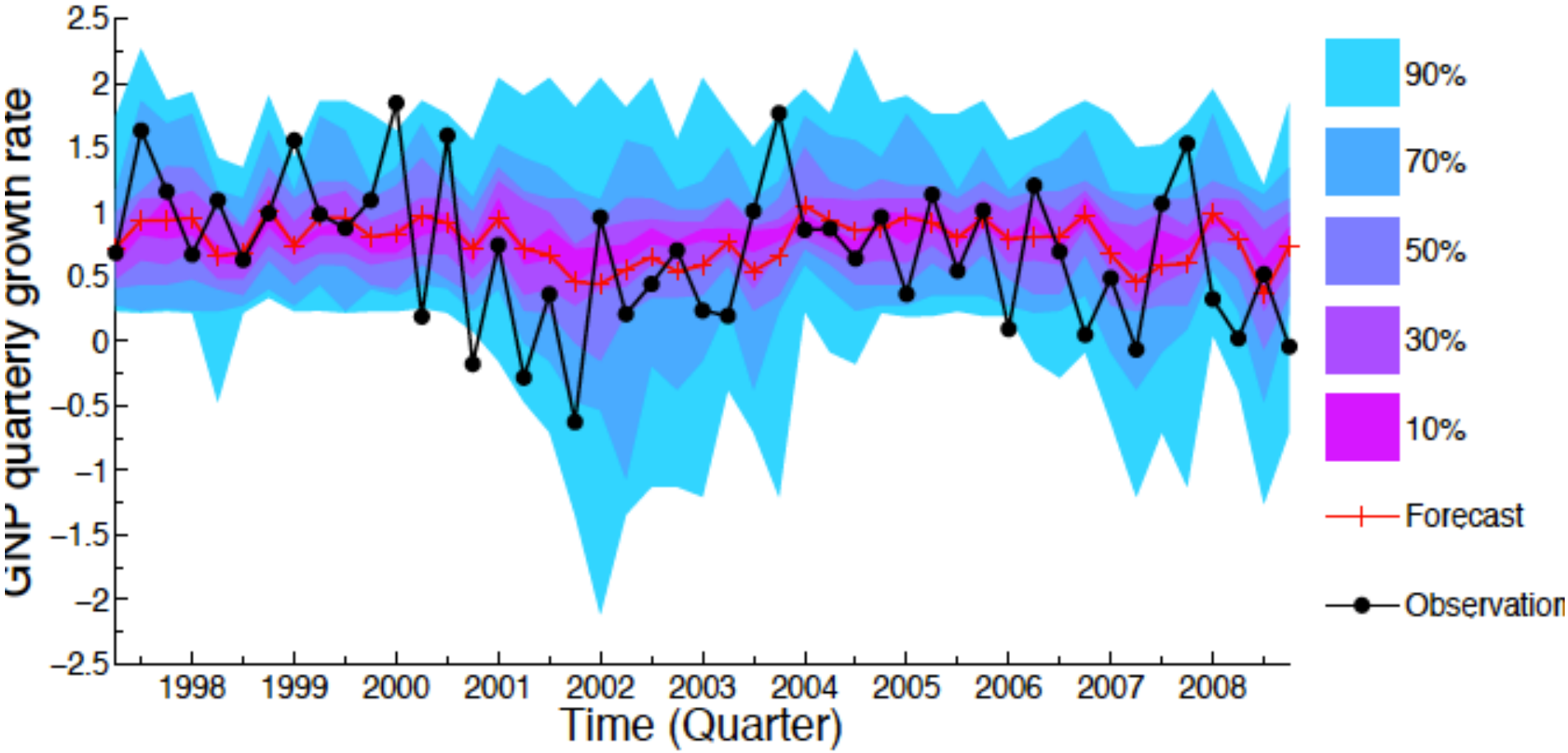
- **Slido.com #66086**

Parameters for nonlinear models

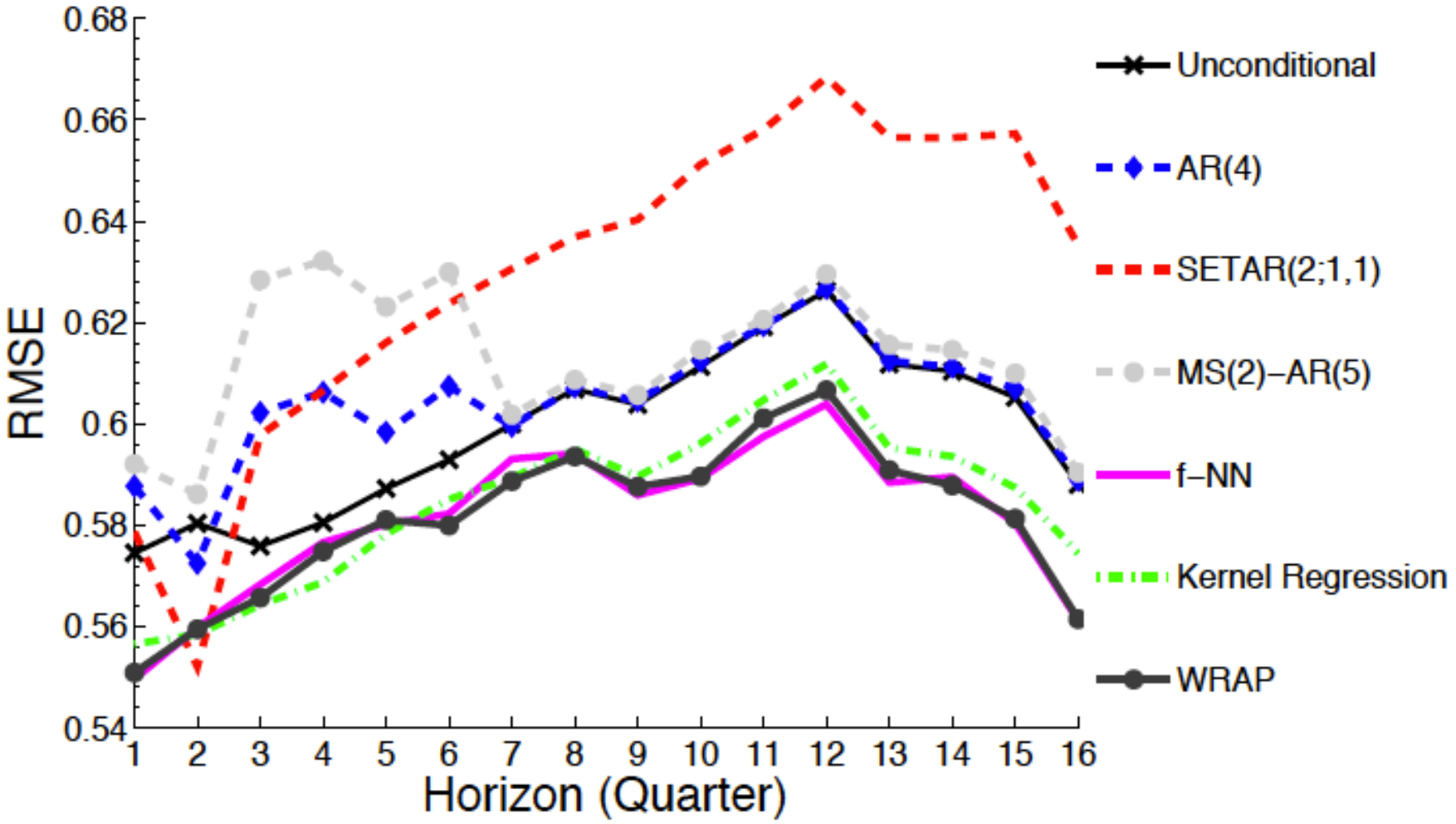
Parameters	SETAR(2;1,1)	MS(2)-AR(5)
Total	8	10

Parameters	F-NN	Kernel Regression	WRAP
f	0.13		0.13
β		0.7	0.12
m	5	5	5
Total	1	1	2

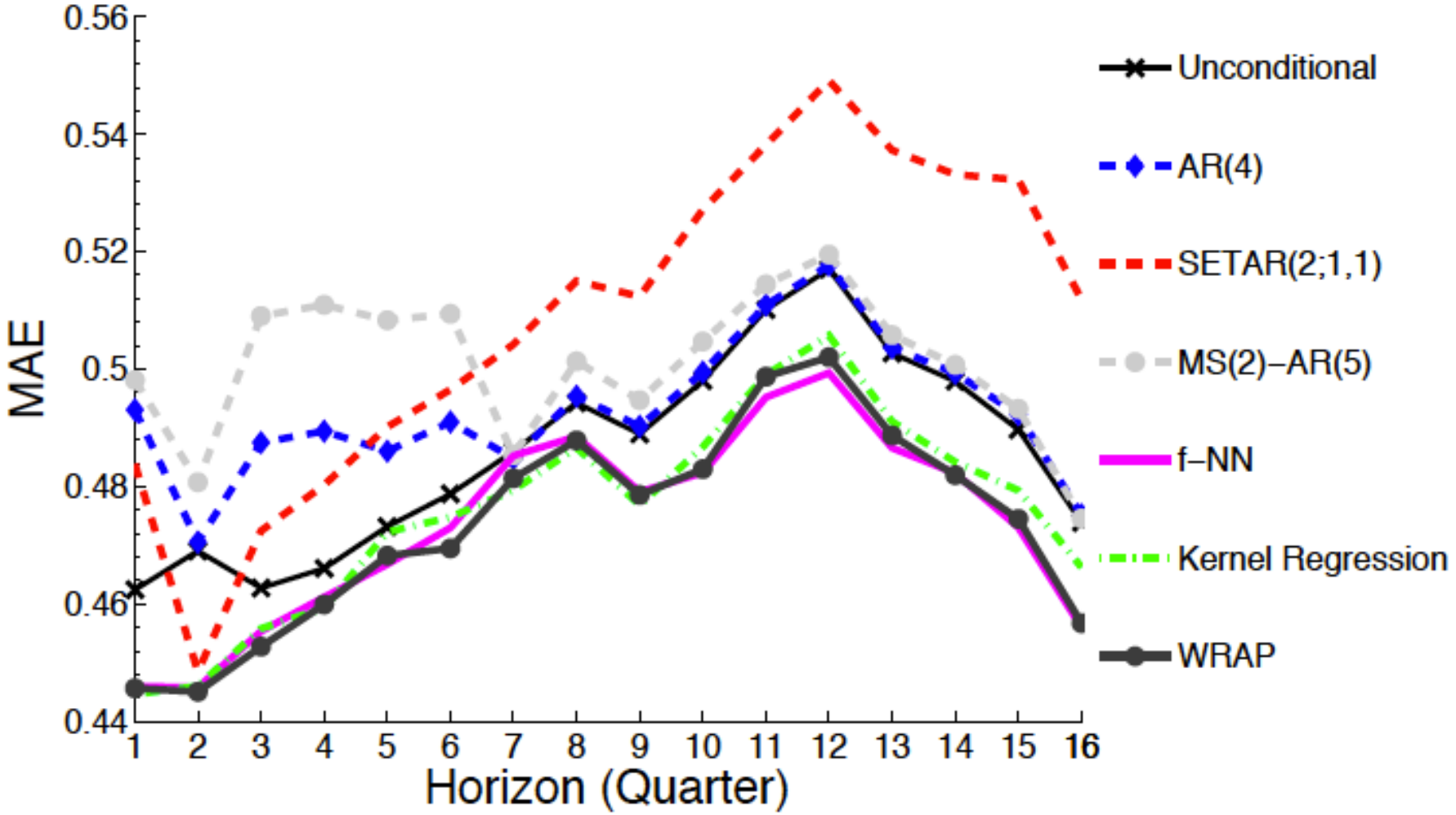
GNP growth rate forecasts



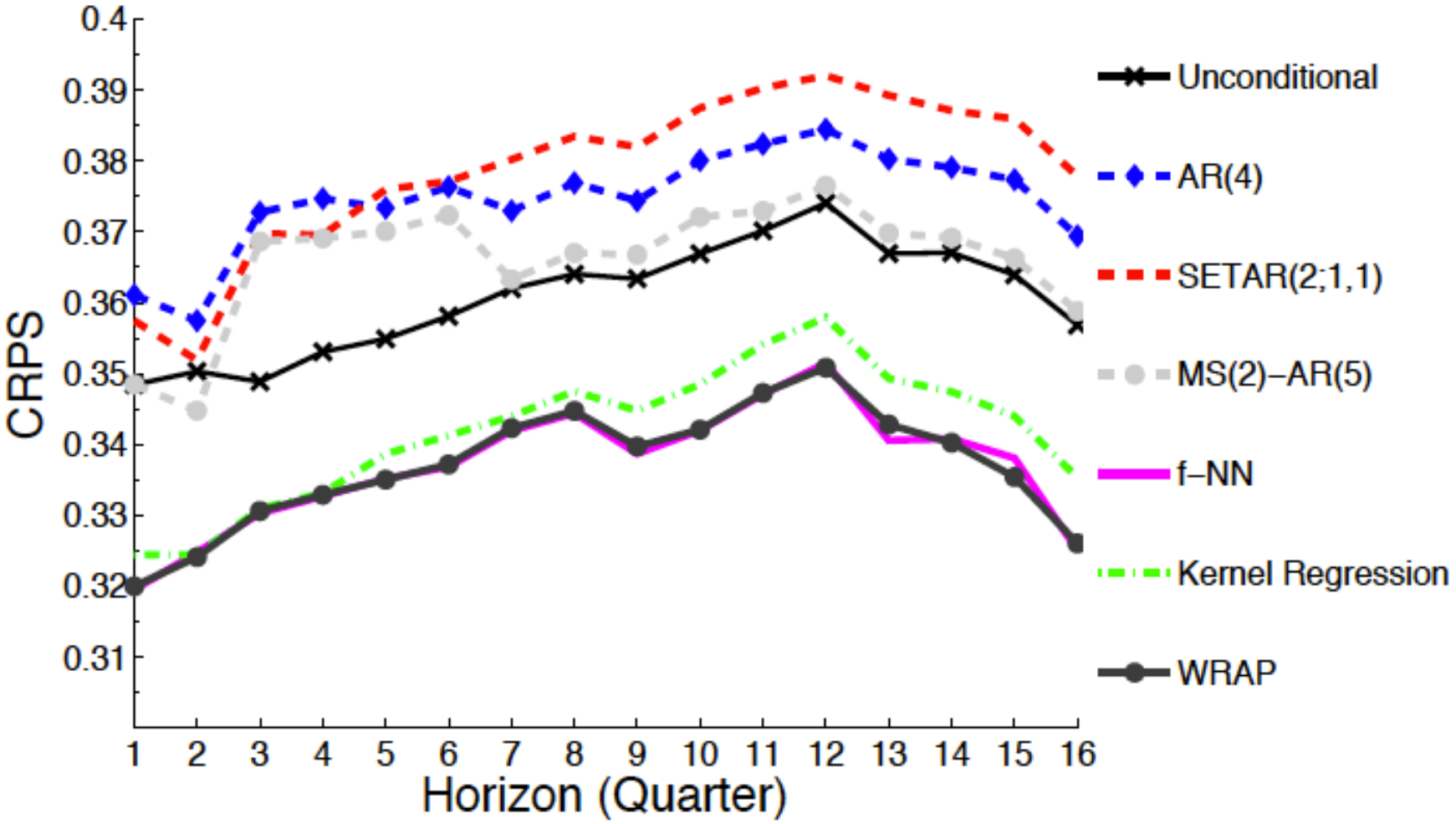
US GNP - RMSE


0.68
0.66
0.64
0.62
0.6
0.58
0.56
0.54
RMSE
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Horizon (Quarter)
Unconditional
AR(4)
SETAR(2;1,1)
MS(2)–AR(5)
f–NN
Kernel Regression
WRAP

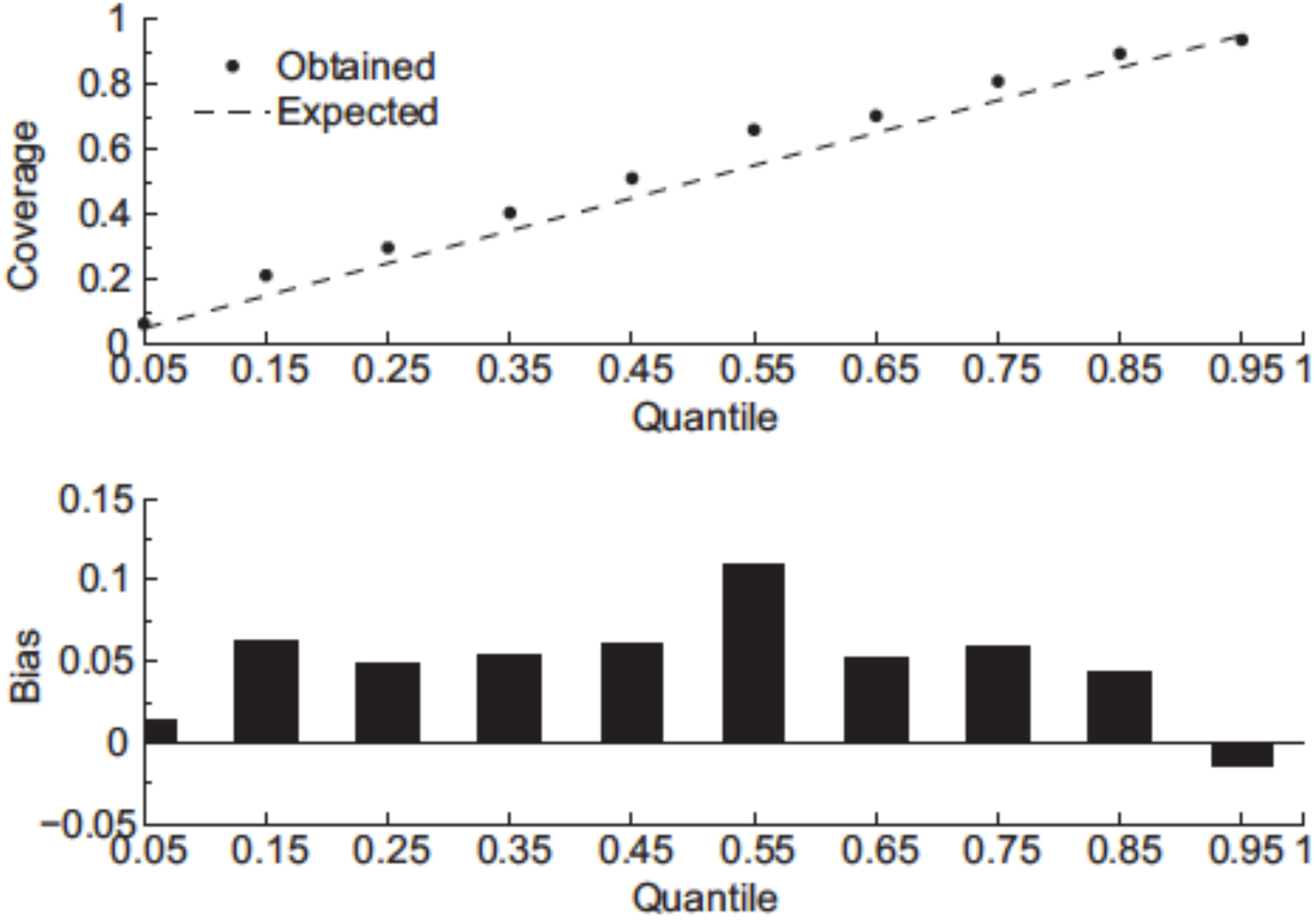
US GNP - MAE


0.56
0.54
0.52
0.5
0.48
0.46
0.44
MAE
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Horizon (Quarter)
Unconditional
AR(4)
SETAR(2;1,1)
MS(2)–AR(5)
f–NN
Kernel Regression
WRAP

US GNP - CRPS


CRPS
0.4
0.39
0.38
0.37
0.36
0.35
0.34
0.33
0.32
0.31
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Horizon (Quarter)
Unconditional
AR(4)
SETAR(2;1,1)
MS(2)–AR(5)
f–NN
Kernel Regression
WRAP

Reliability



Summary of Results

- MSE-F tests: Parametric models (AR, SETAR and MS-AR) do not provide any additional information over the unconditional benchmark, whereas the nonparametric models (*f*-NN, kernel regression and WRAP) offer improvement in forecast accuracy at all horizons
- KS tests: Forecasts from nonparametric models are significantly different from the parametric models
- PIT: nonparametric models are relatively more uniformly distributed compared to parametric models

Conclusions

- Unconditional mean (long-run average) is more competitive than persistence benchmark
- A parsimonious, nonlinear and nonparametric model that generates accurate point and density forecasts of US GNP
- Advantages are clearer for density forecasts
- Outperforms previously published nonlinear models with several parameters
- Adaptive approach for data streams

Arora, Little & McSharry (2013), Studies in Nonlinear Dynamics & Econometrics (in press).