ITMO UNIVERSITY

Saint Petersburg, Russia

Methods and models for multivariate data analysis

Lecture 2. A bit about time (practice).

Dr. Alex Hvatov, Composite AI Lab Head

SPb

29.11.2022

Fac et excusa– act now, excuse later

(Finite difference processes)

ITMO UNIVERSITY

Main modelling concepts:

- What if we can make a stationary process and study it?
- How do we model the non-stationary part?
  - Polynomial Fitting
  - AR* models
  - Fourier transform (?)
  - DE
  - …
  - And what is the difference between $T(t)$ and $S(t)$? Should we remove $S(t)$?

ITsMOre than a UNIVERSITY

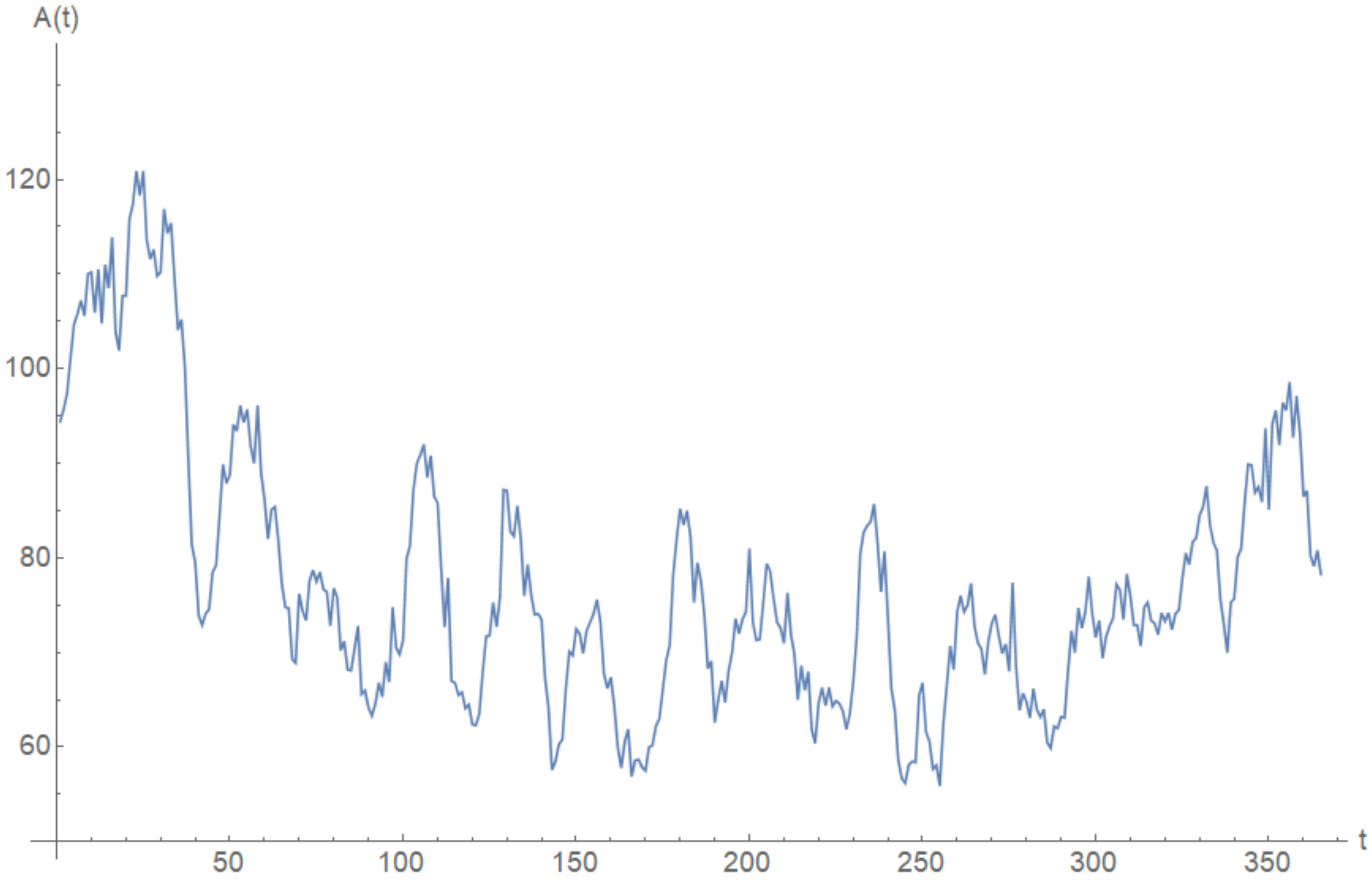
Flashback: AR(p) models

- Model is: $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots a_p y_{t-p} + \varepsilon_p$
  - $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots a_p y_{t-p}$ - with truncated noise
  - We need something like a linear system to find a_i
  - $A = \begin{pmatrix} (y_{t-1}, y_{t-1}) & \cdots & (y_{t-1}, y_{t-p}) \\ \vdots & \ddots & \vdots \\ (y_{t-1}, y_{t-p}) & \cdots & (y_{t-p}, y_{t-p}) \end{pmatrix}$,b $= \begin{pmatrix} (y_t, y_{t-1}) \\ \vdots \\ (y_t, y_{t-p}) \end{pmatrix}$
  - Solve for x: $Ax = b$, where $x = (a_1, \dots a_p)$
  - After that unit root test is used

Trend T(t) and I,pt.I

- Model is: $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots a_4 y_{t-4}$
- Coefficients:{0.3815,0.3815,0.4416,0.9987}

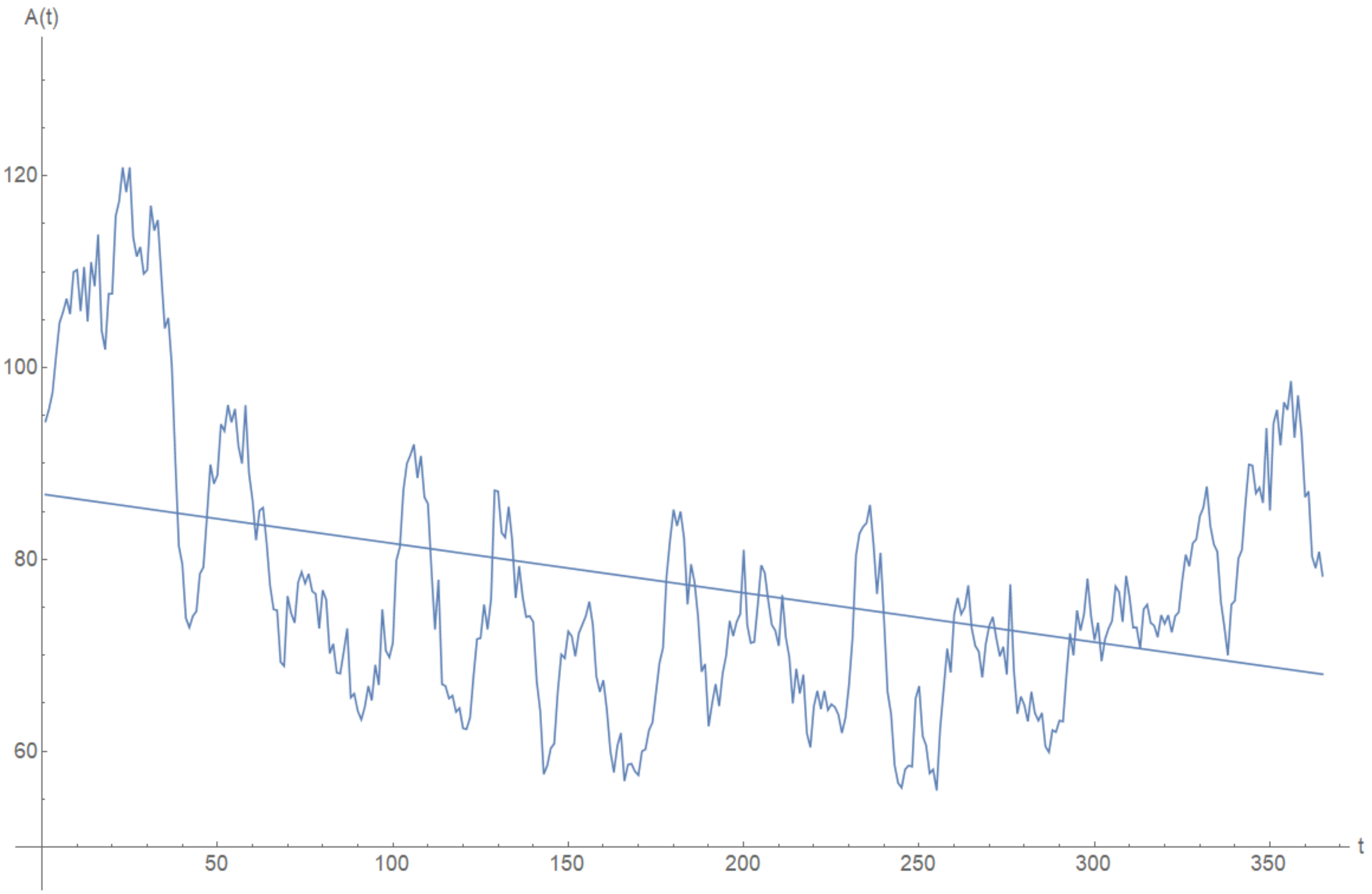
  Just one root close to one. So, a linear trend?



Trend T(t) and I,pt.II

- Model is: $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots a_4 y_{t-4}$
- Coefficients:{0.3815,0.3815,0.4416,0.9987}

  Just one root close to one. So a linear trend?



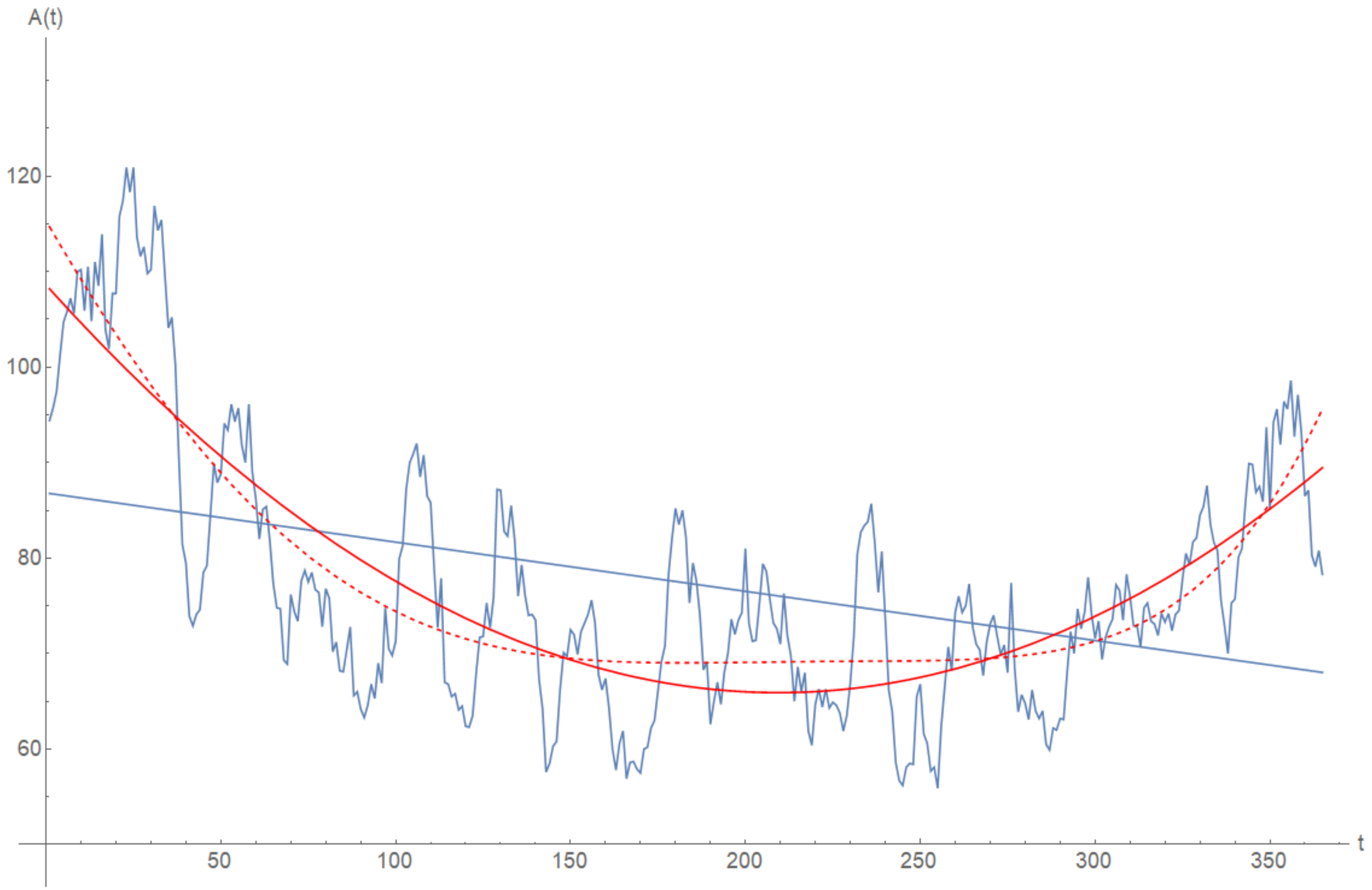
Trend T(t) and I,pt.II

- Model is: $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots a_4 y_{t-4}$
- Coefficients:{0.3815,0.3815,0.4416,0.9987}

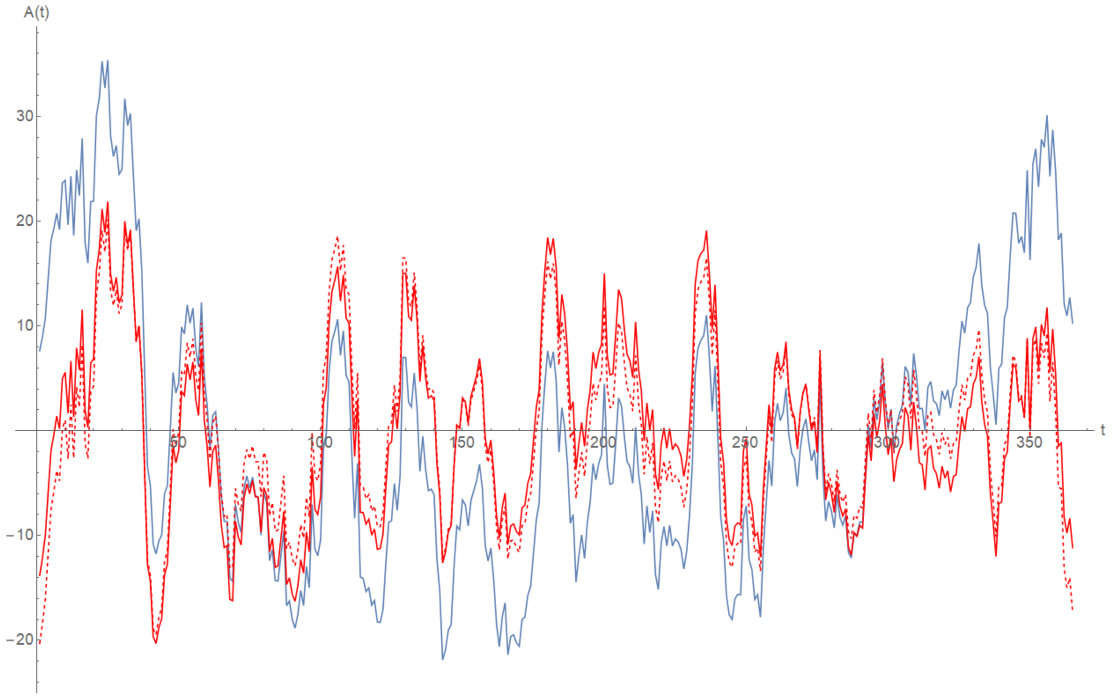
  Just one root close to one. So a linear trend?

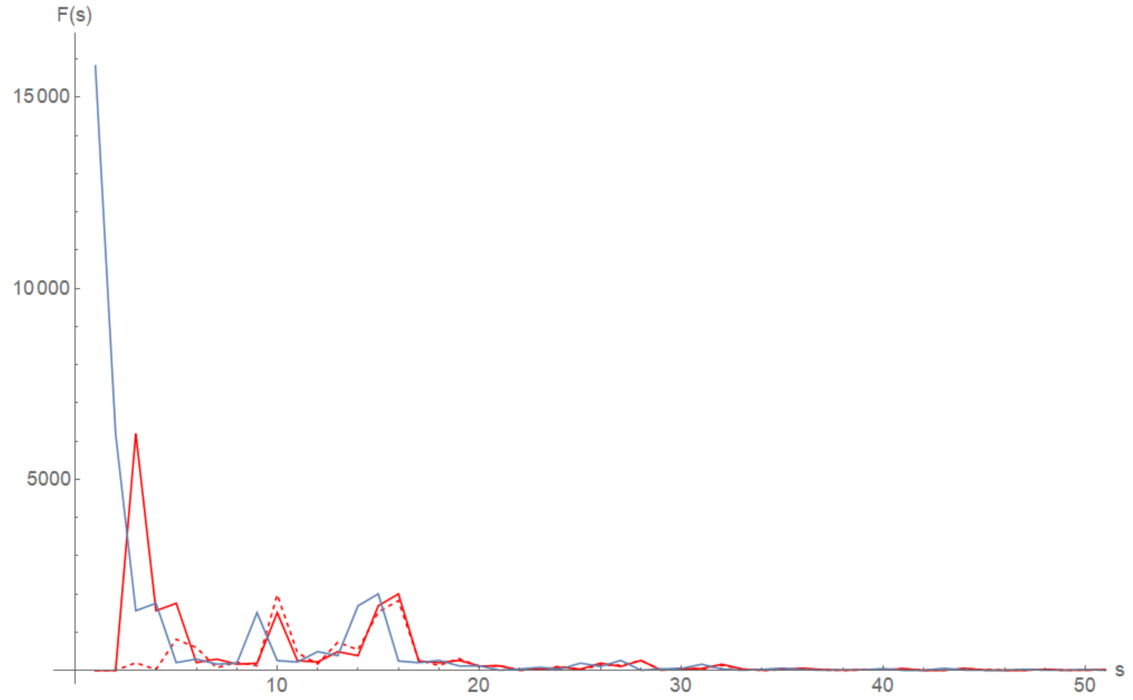
  Parabola?

  4^{th} degree polynomial?



Trend T(t) and I,pt.3





- Residual time series
- Specter of residuals
- Specter is <u>shifting</u> when trend is not removed!

Reminder: Fourier Series

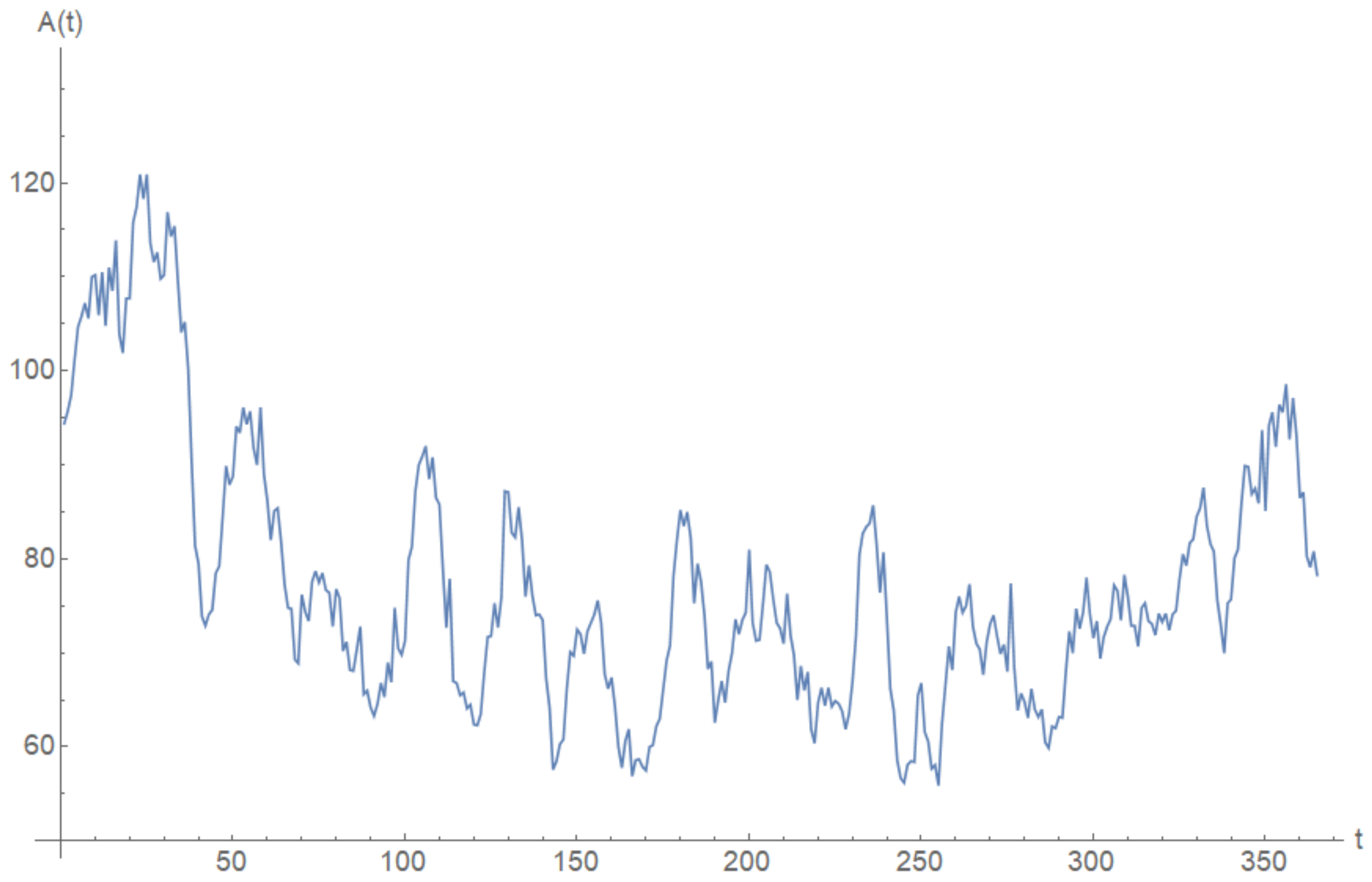
- $f(t) = \frac{a_0}{2} + \sum_{s=1}^{n} a_s \cos\frac{2\,\pi\,s}{n}t + b_s \sin\frac{2\,\pi\,s}{n}t$ - Discrete Fourier series
  - $a_s = \sum_{k=1}^{n} f(k) \cos\frac{2\,\pi\,s}{n}k$
  - $b_s = \sum_{k=1}^{n} f(k) \sin\frac{2\,\pi\,s}{n}k$

- Bugcat Capoo wants you to remember Fourier transform, it will be used today

$$F(s) = \int_{-\infty}^{\infty} f(t)\exp(-2\pi st)dt$$

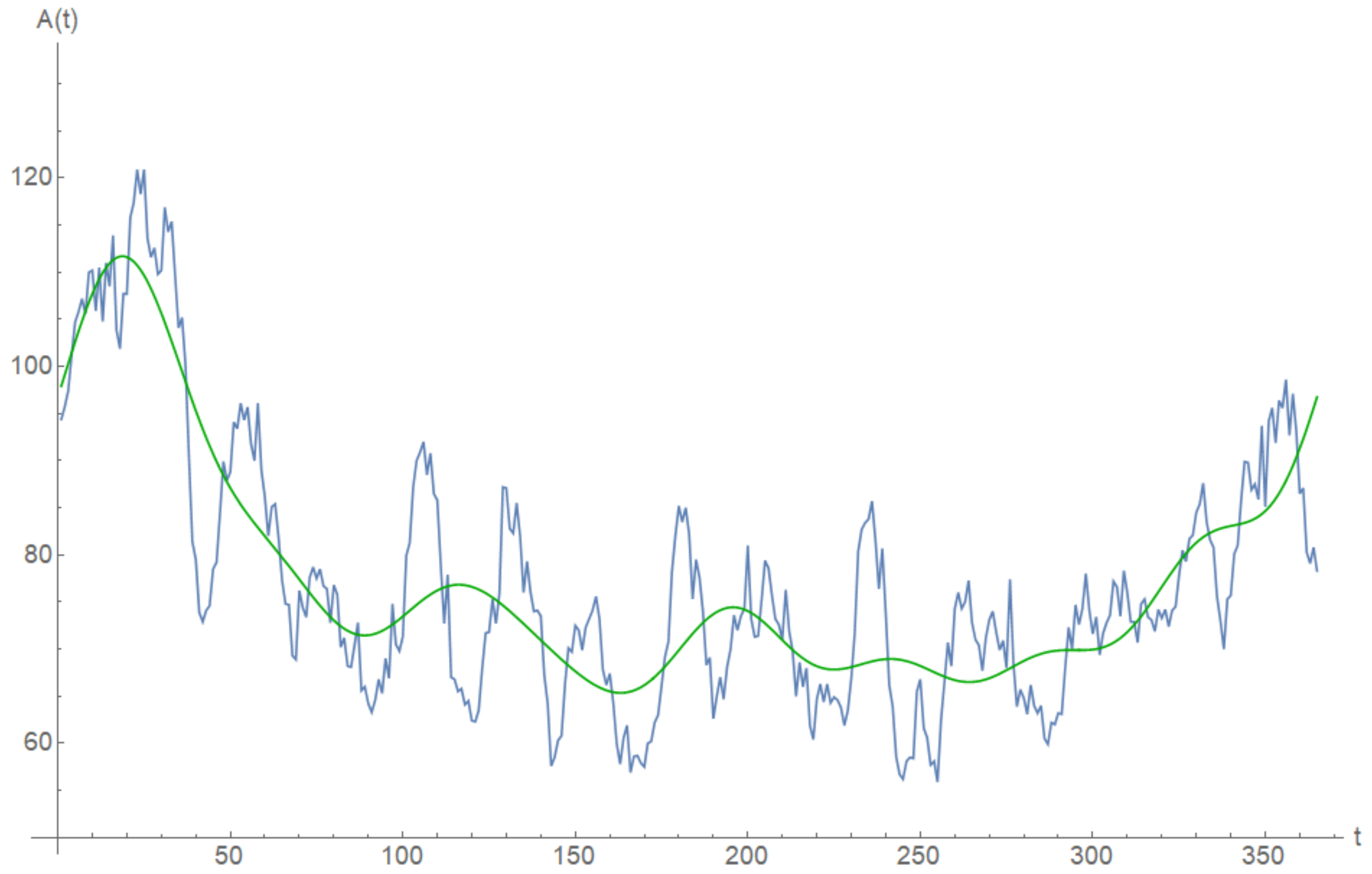
Seasonal component S(t) and I,pt.I

- What if we will ignore polynomial part and start with seasonal?
- What is the physical meaning of zero frequency?
- What is the length of the season?



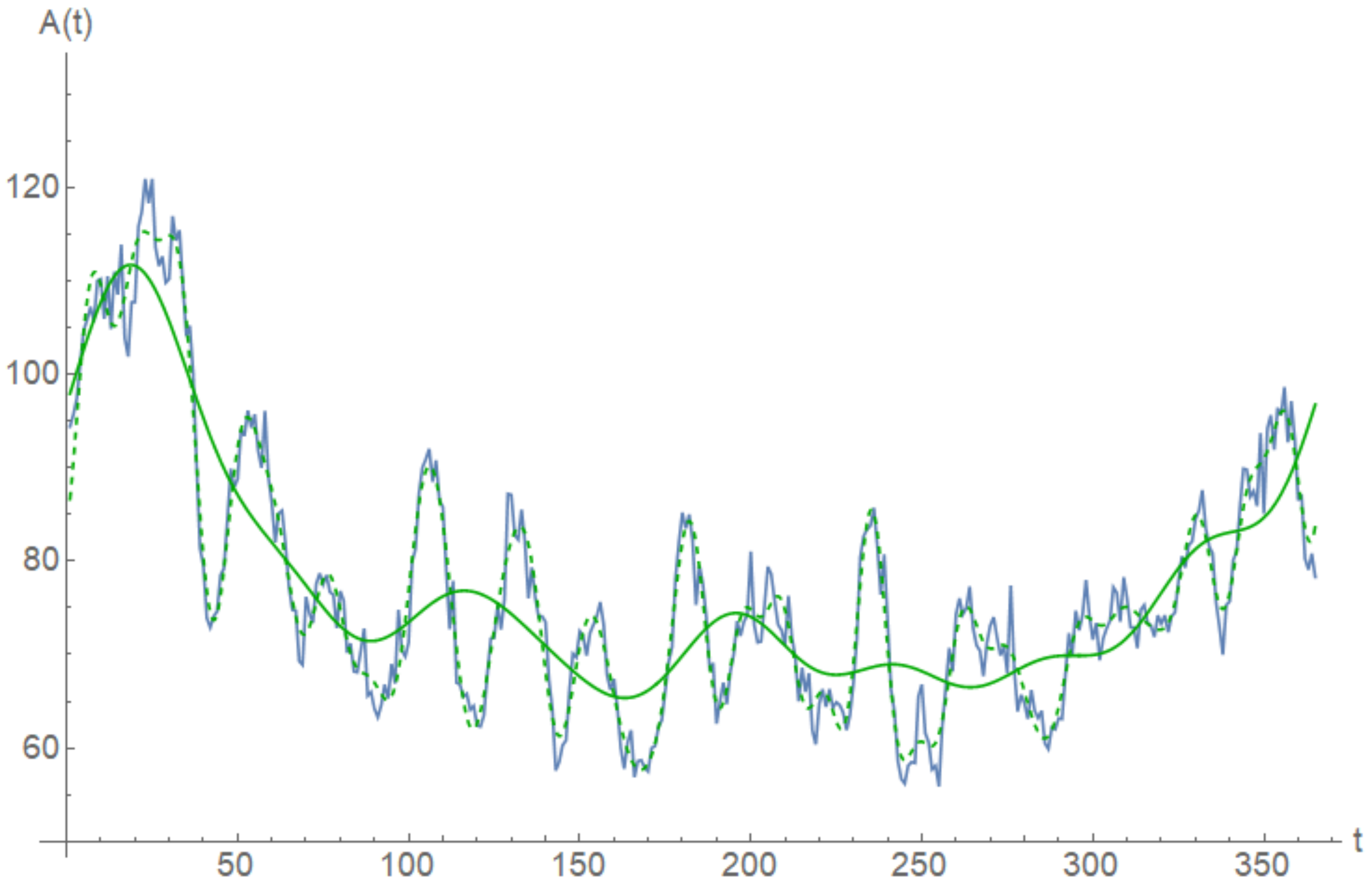
Seasonal component S(t) and I,pt.I

- What if we will ignore polynomial part and start with seasonal?
- What is the physical meaning of zero frequency?
- What is the length of the season?
- 8 days?

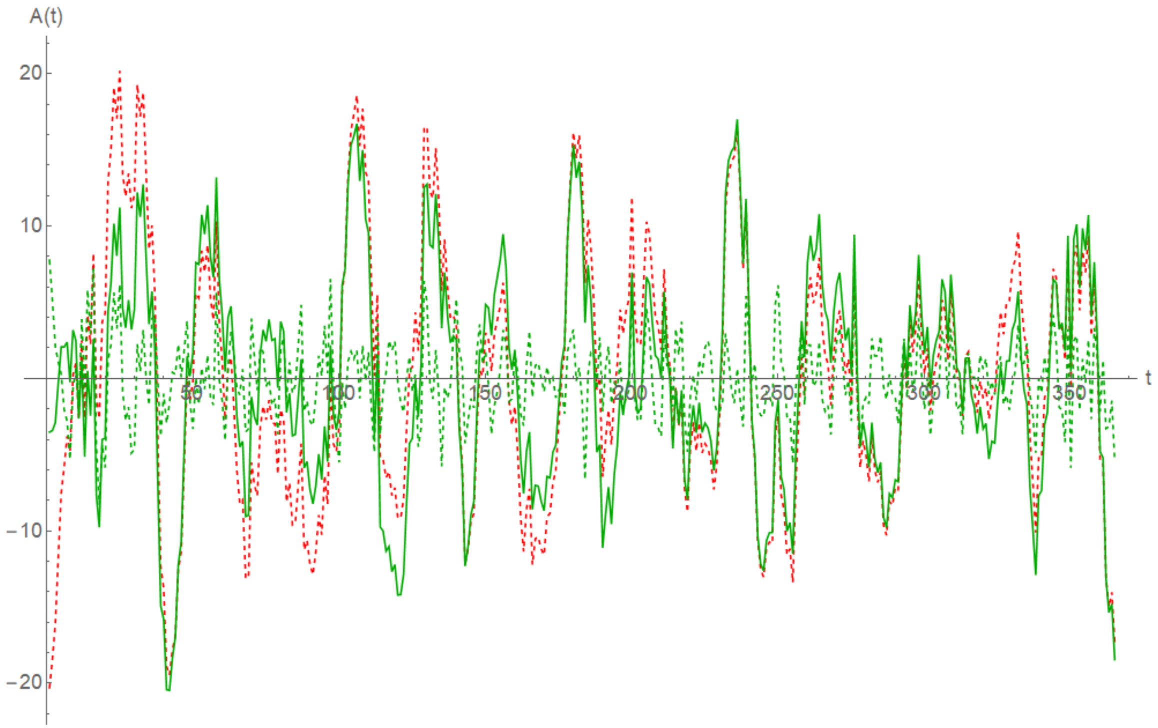


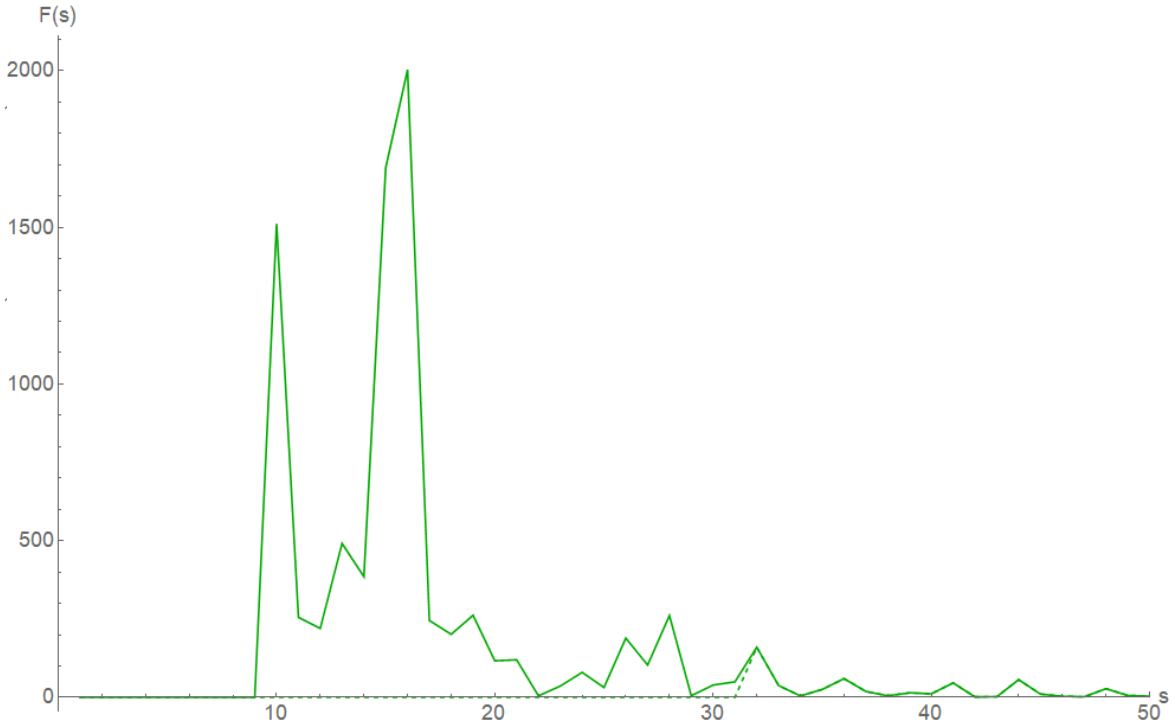
Seasonal component S(t) and I,pt.I

- What if we will ignore polynomial part and start with seasonal?
- What is the physical meaning of zero frequency?
- What is the length of the season?
- 30 days?



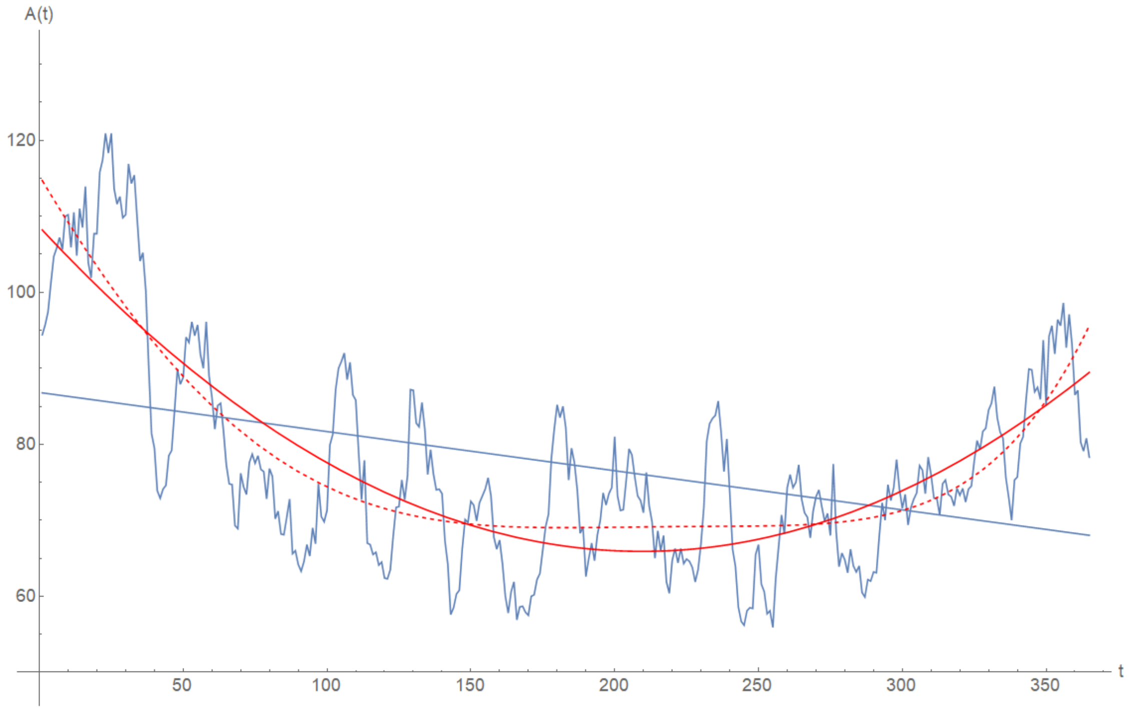
Seasonal component S(t) and I,pt.II

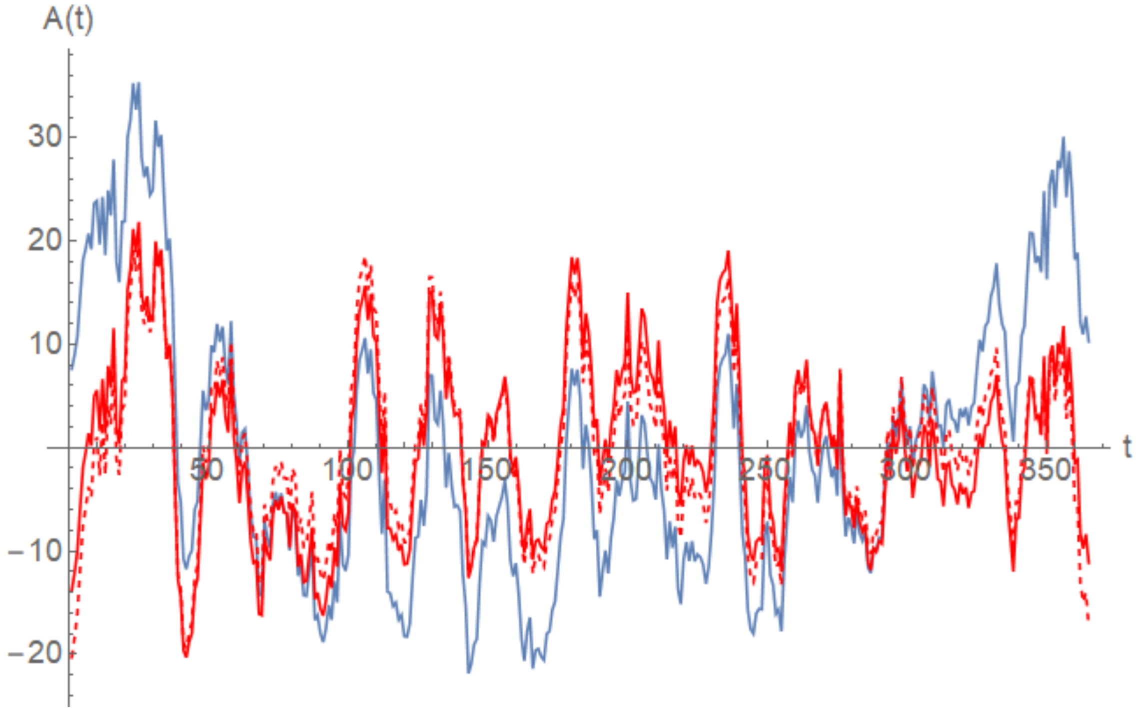




- Residual time series
- Specter of residuals
- Low frequencies are removed with seasonal component – we lose information!

The (classical) plan - I





- Remove polynomial trend T(t)
- Good place to stop R2>0.5

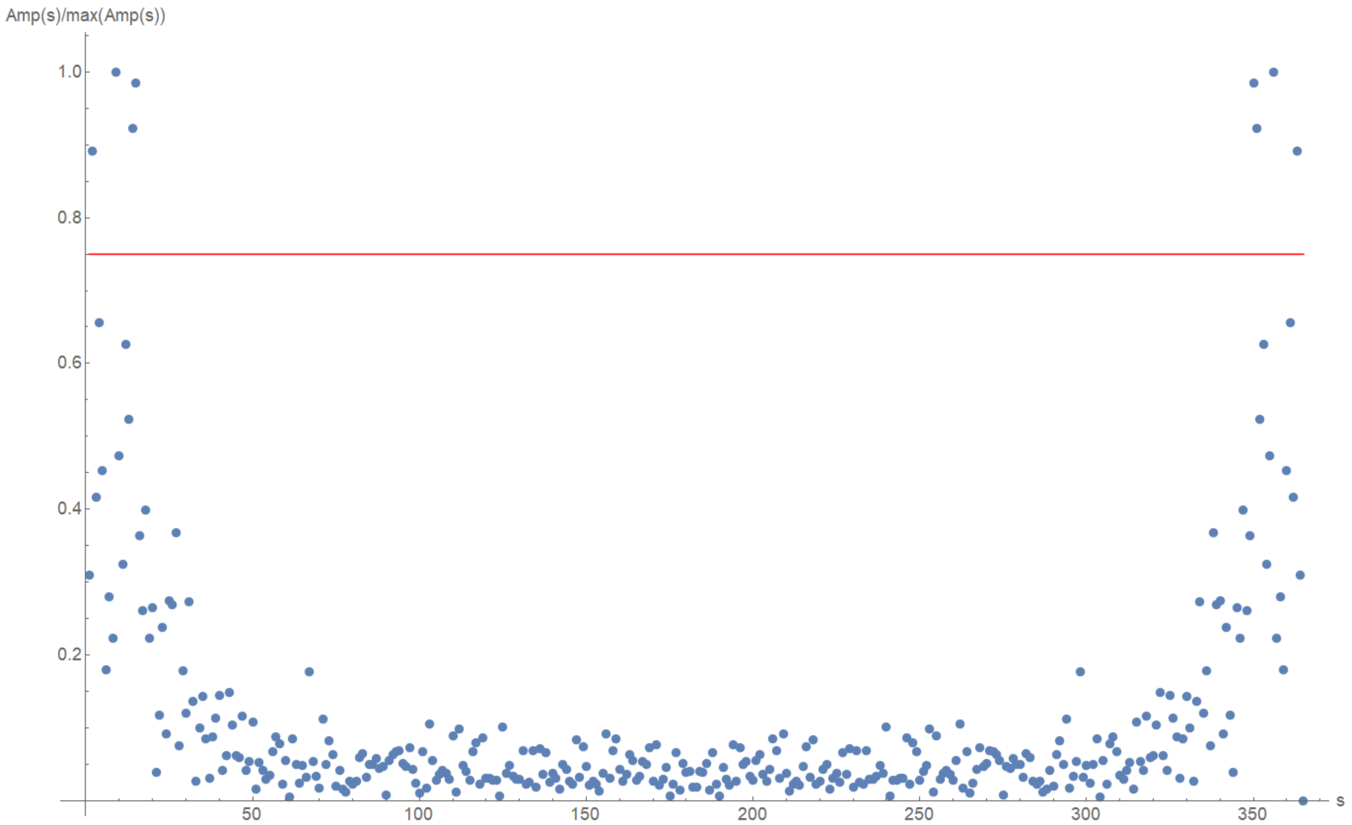
Linear: R2=0.15

<u>Quadric: R2=0.62</u>

4th degree: R2=0.66

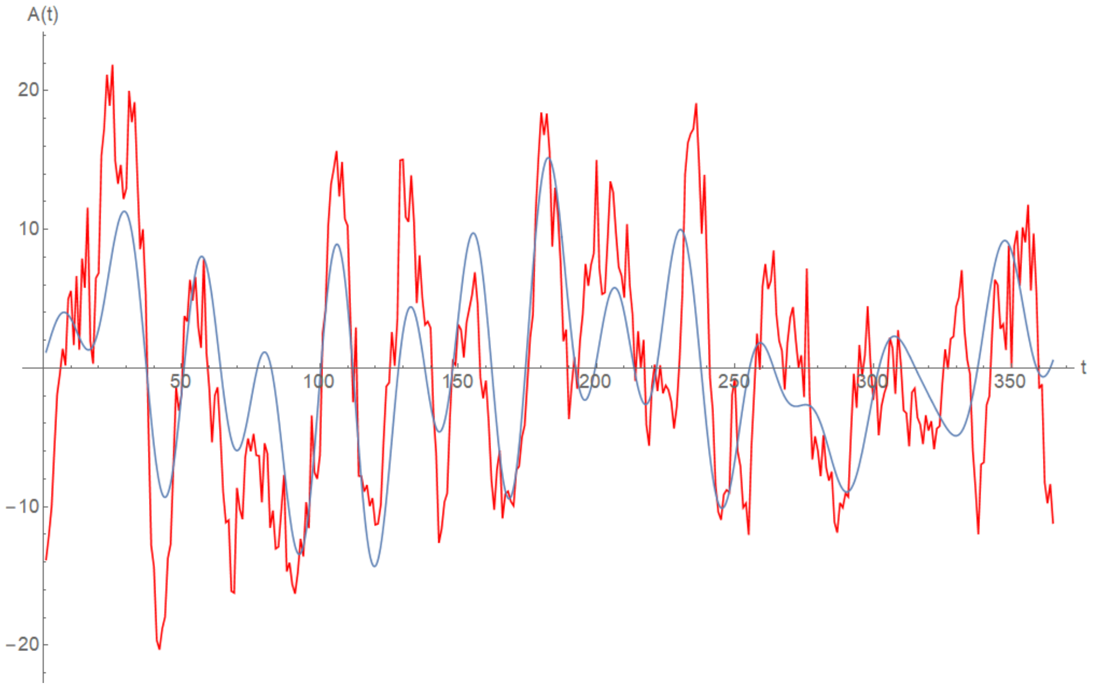
The (classical) plan - II

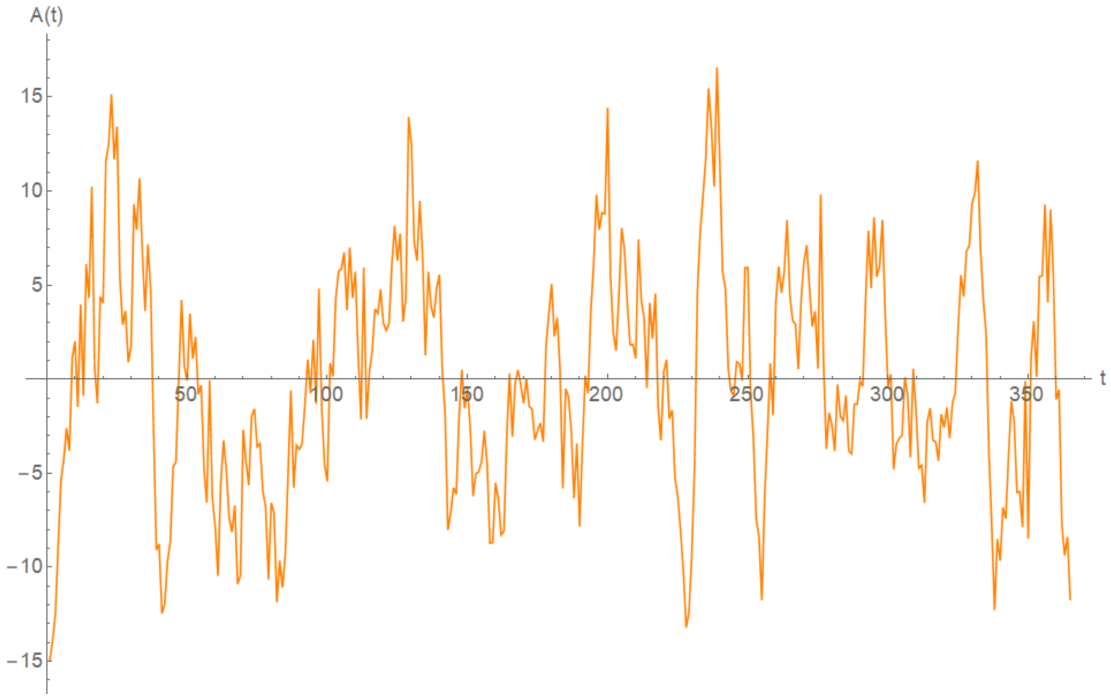
- Seasonal part is more tricky, since it affects specter
- Good place to start is to plot amplitudes $\sqrt{a_s^2 + b_s^2}$
- More informative $\frac{\sqrt{a_s^2+b_s^2}}{\max\sqrt{a_s^2+b_s^2}}$
- Seasonal component is waves with frequency higher than ~ 0.75



Seasonal frequencies here 2,9,14,15

The (classical) plan - III





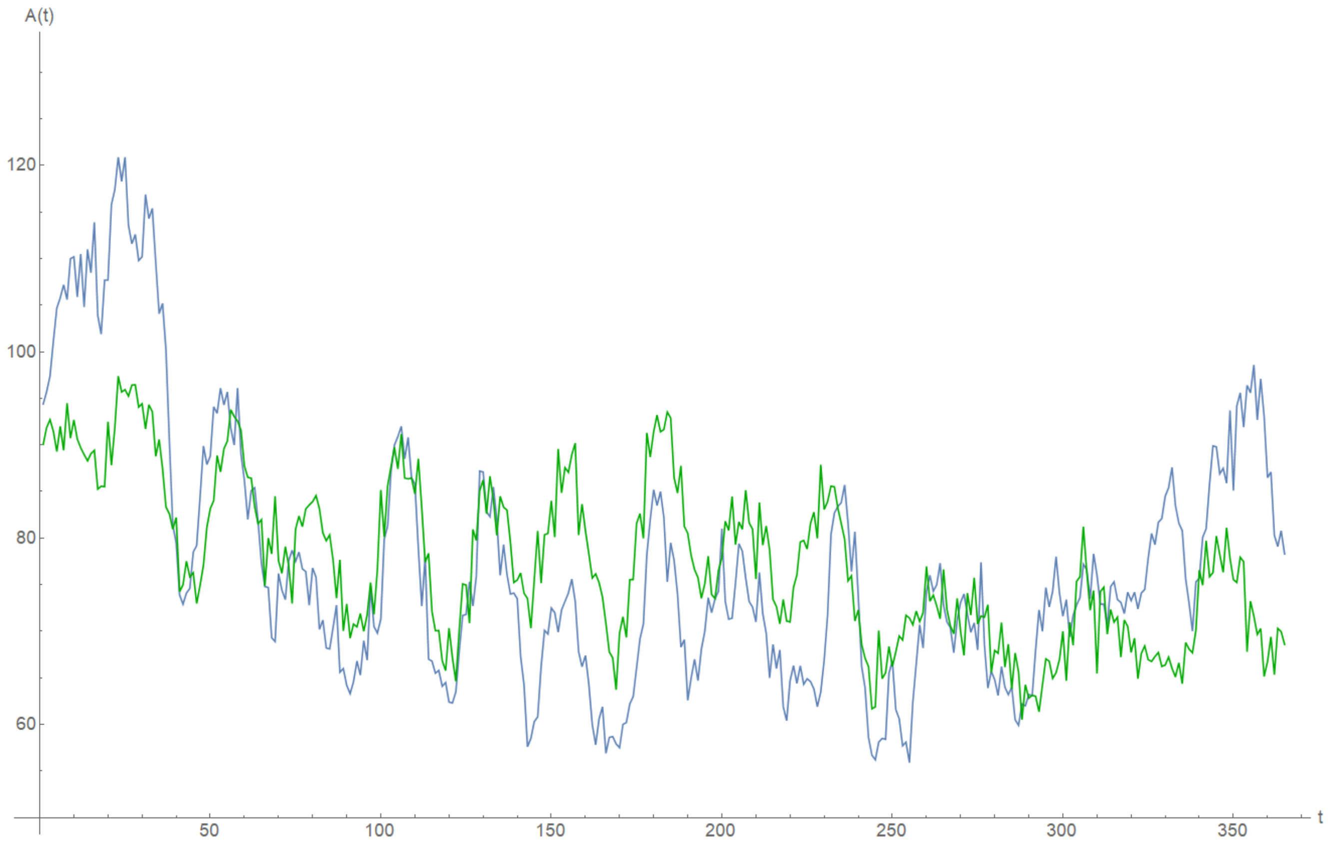
- Remove seasonal component S(t)
- Model noise ε_t in usual way as stationary process
- Total model is f(t)=T(t)+S(t)+ε_t

$$R2 = 0.814$$
$$m_t \sim 10^{-13}$$
$$D_t \sim 4.867$$

ITMOre than a UNIVERSITY

The (classical) plan - final

- Well, mean error is 8.58
- Mean value is 77.4
- Pretty good result for 10 minutes of approximation
- Better work with noise – better results



The (modern) plan

- What if we still can make a stationary process and study it?
- How do we model the non-stationary part?
  - LSTM-like
  - Prophet (I do not want to think way)
  - Various esoteric methods (algebraic decomposition, neural-ODE)
- $f(t) = F(t) + \varepsilon(t) = G(t)\varepsilon'(t)$

LSTM way (I sign up to train NNs)

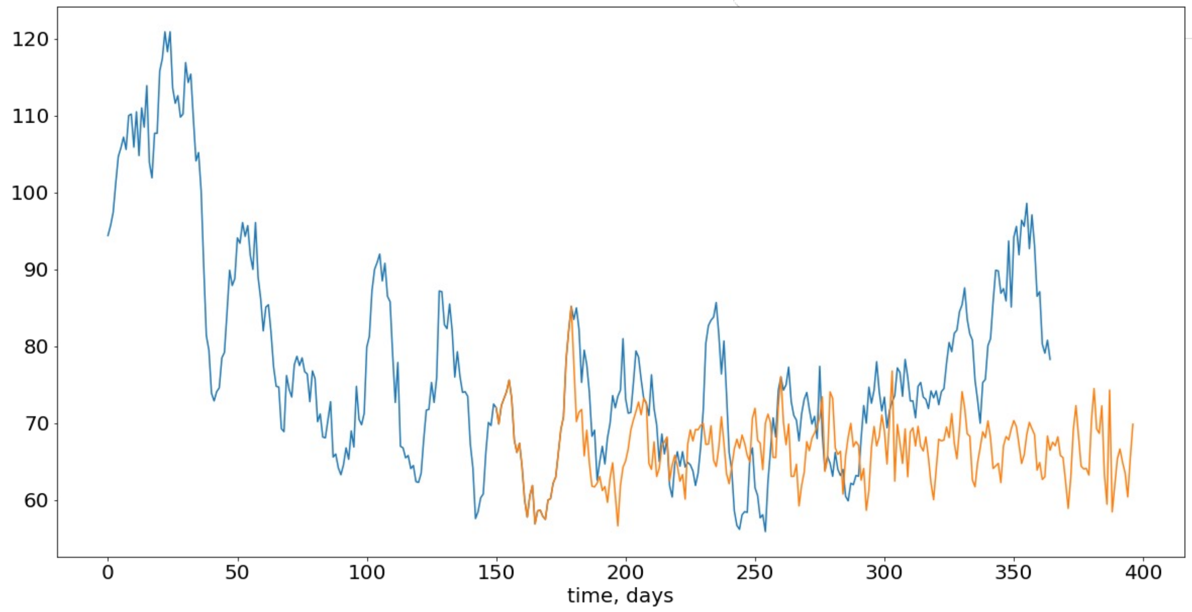
ITMO UNIVERSITY

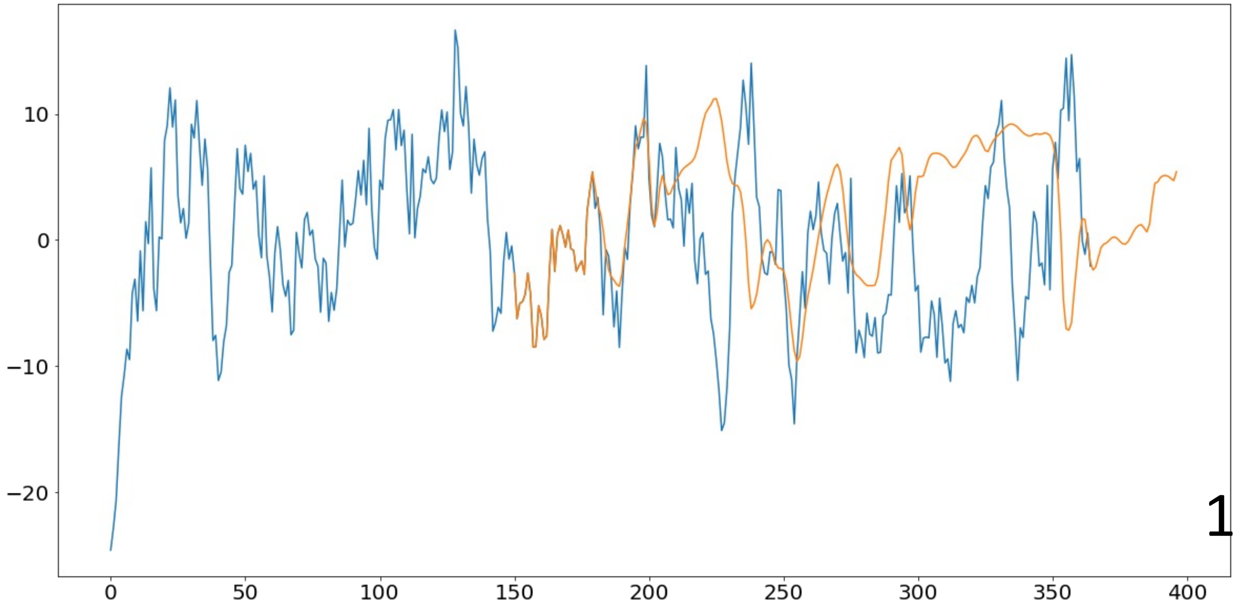
- Classical LSTM Model may be written as:
  $a_{t+M}y_{t+M} + \cdots + a_{t+1}y_{t+1} + a_t y_t = a_{t-1}y_{t-1} + a_{t-2}y_{t-2} + \cdots a_{t-N}y_{t-N} + \varepsilon_p$ (whoa (!) it looks like autoregression)
- Trend shifts specter issue
- Noised in -> noised out

ITsMOre than a UNIVERSITY

LSTM way (I sign up to train NNs)

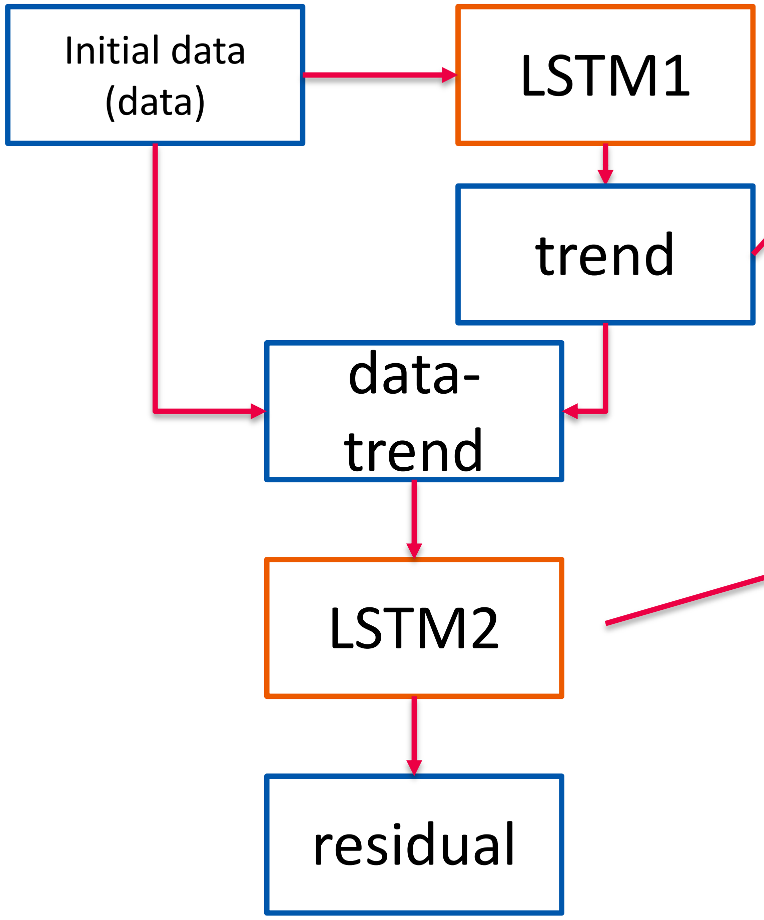
- Classical LSTM Model predicts only stationary processes
- Maybe not so good
- We could also use the several LSTMs



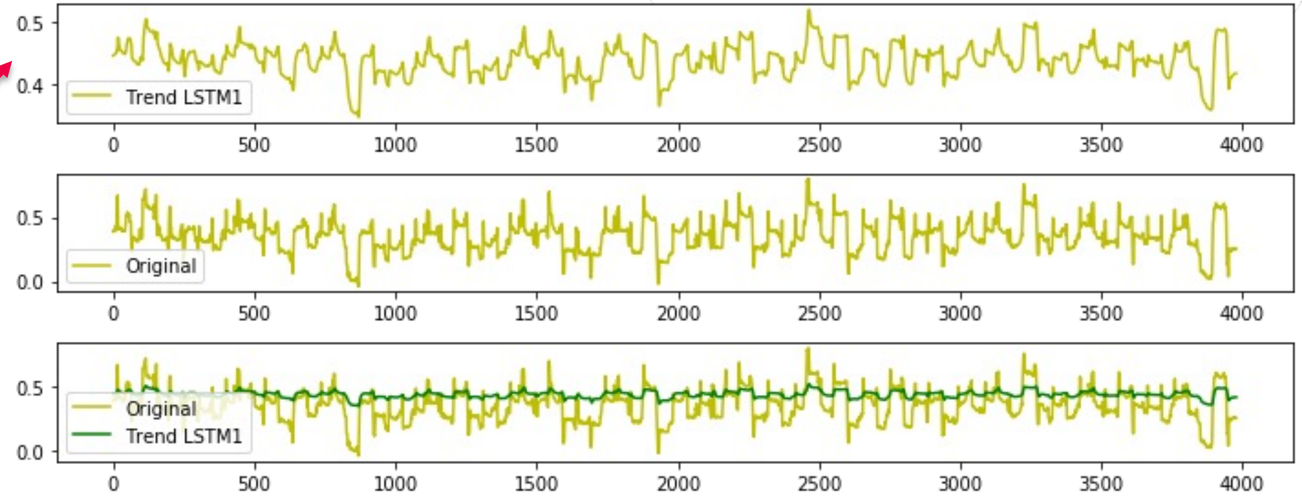


LSTMs way (I want to be the NNs pro)

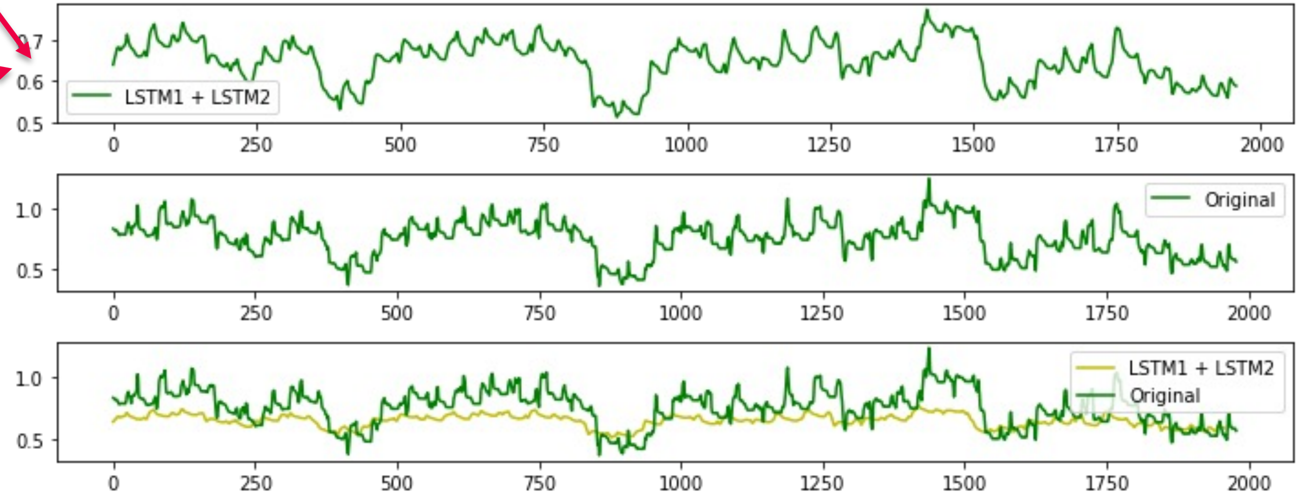
- Two LSTMs allow us to represent two scales



LSTM scheme



Trend LSTM example

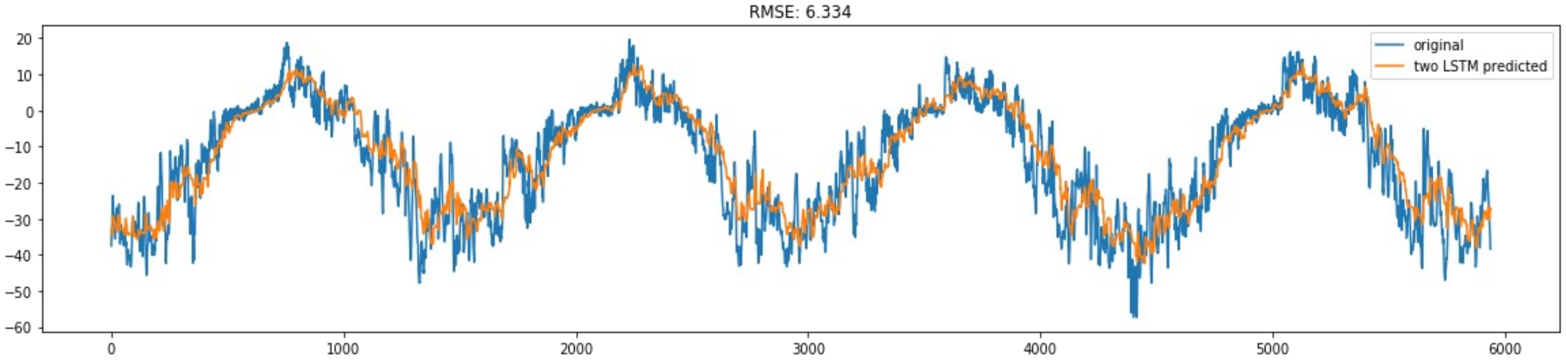


Total LSTM example

LSTMs way (I want to be the NNs pro)

Non-stationary part:

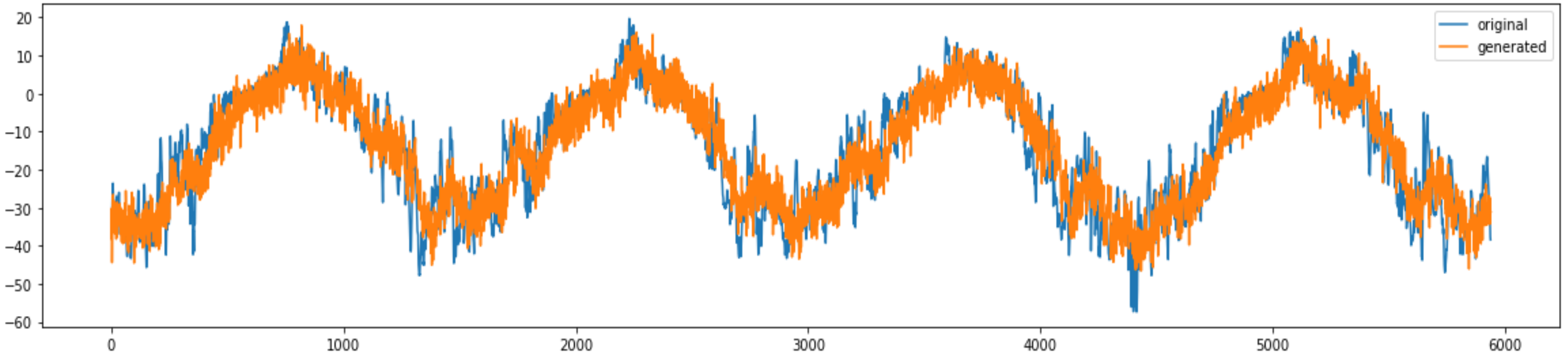
- Not really differs from trend-season decomposition part



Non-stationary part

Noise part:

- We add noise to the predicted nonstationary part

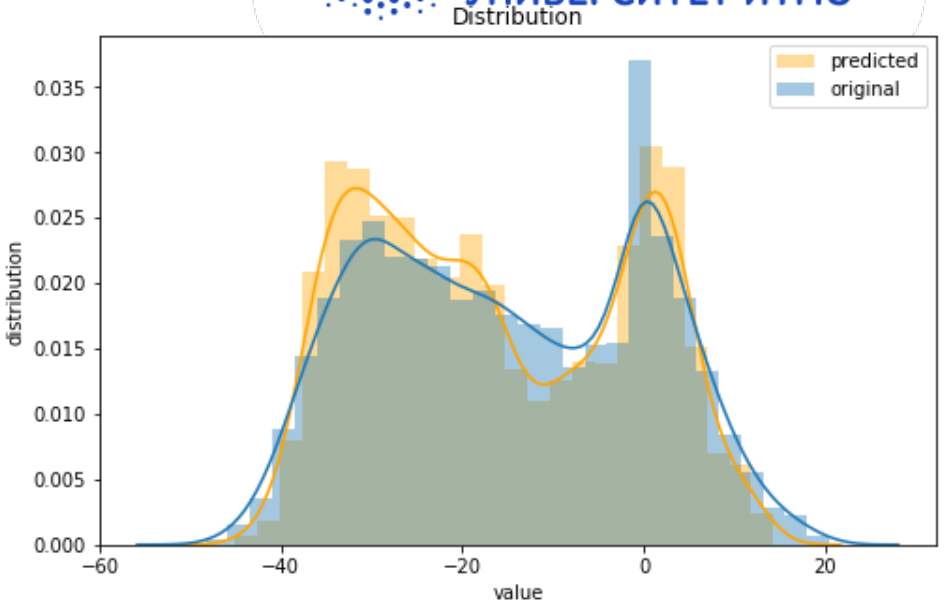


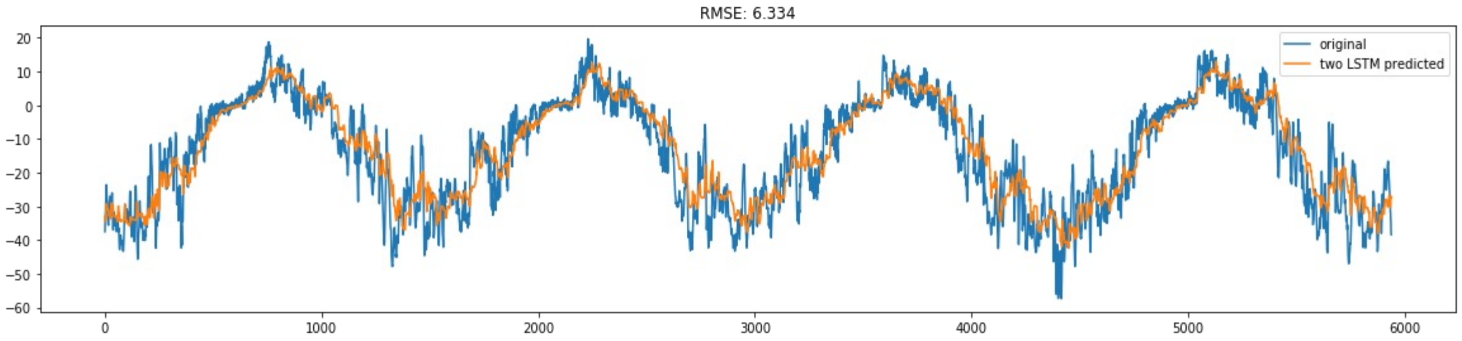
Non-stationary part+noise

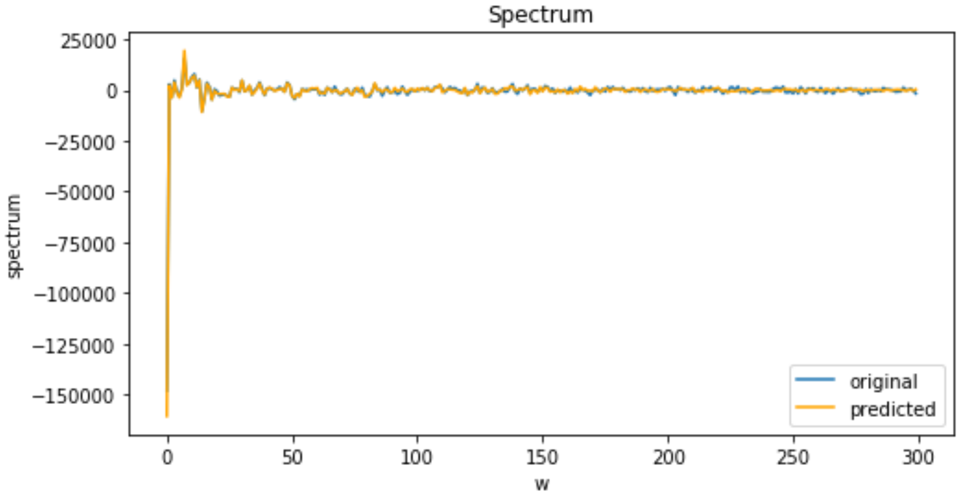
LSTMs way (I want to be the NNs pro)

How do we assess modelling:

- RMSE (it is a stochastic process though)
- Distribution
- Spectre



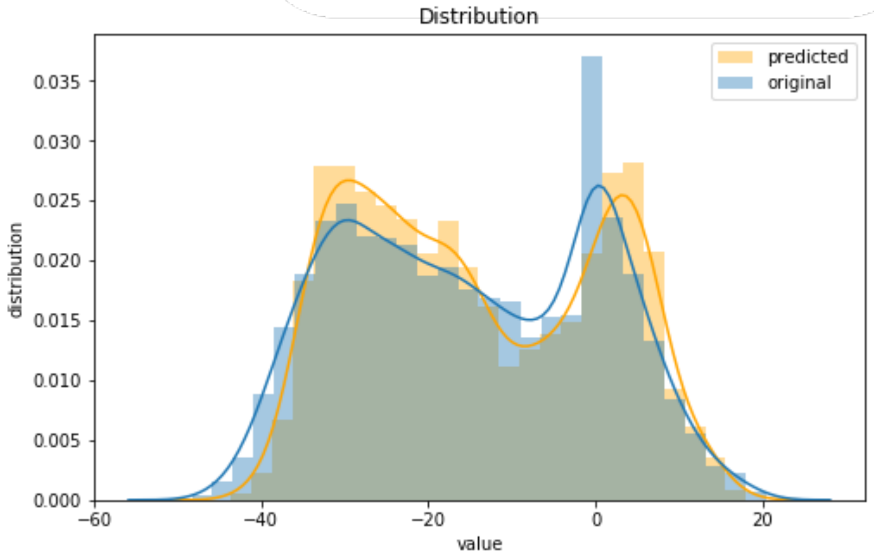


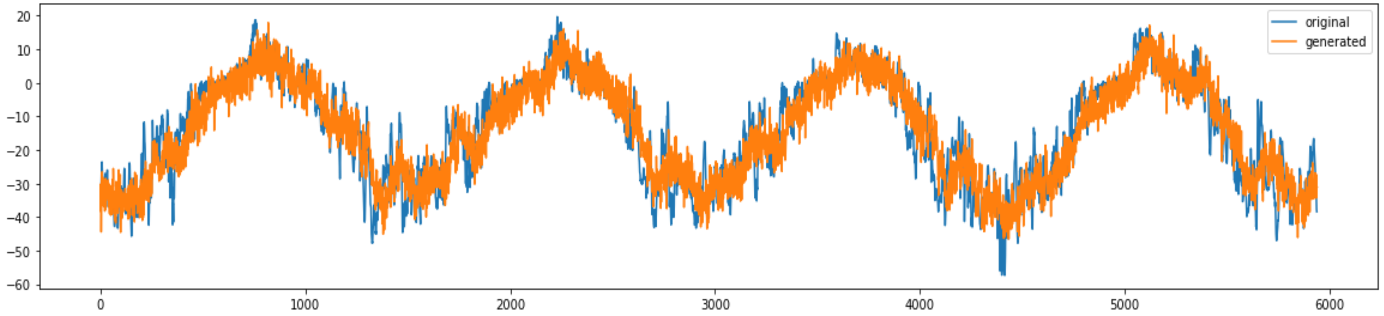


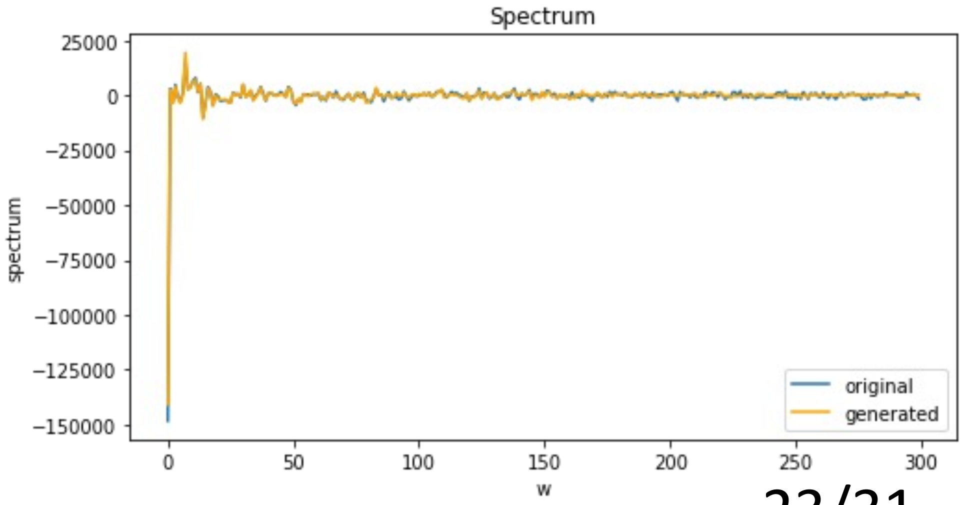
LSTMs way (I want to be the NNs pro)

How do we assess modelling:

- RMSE (it is a stochastic process though)
- Distribution
- Spectre

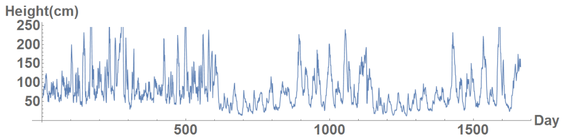






Prophet (I do not want to think way)

Time series



trend season holidays

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

noise

$$g(t) = \frac{C(t)}{1 + \exp(-(k + a(t)^T\delta)(t - (m + a(t)^T\gamma)))} \sim \frac{C}{\exp(-k(t-m))}$$ for non-stationary TS

$g(t) = (k + a(t)^T\delta)(t - (m + a(t)^T\gamma))$ – for close to stationary time-series

$s(t)$ - season as ususal

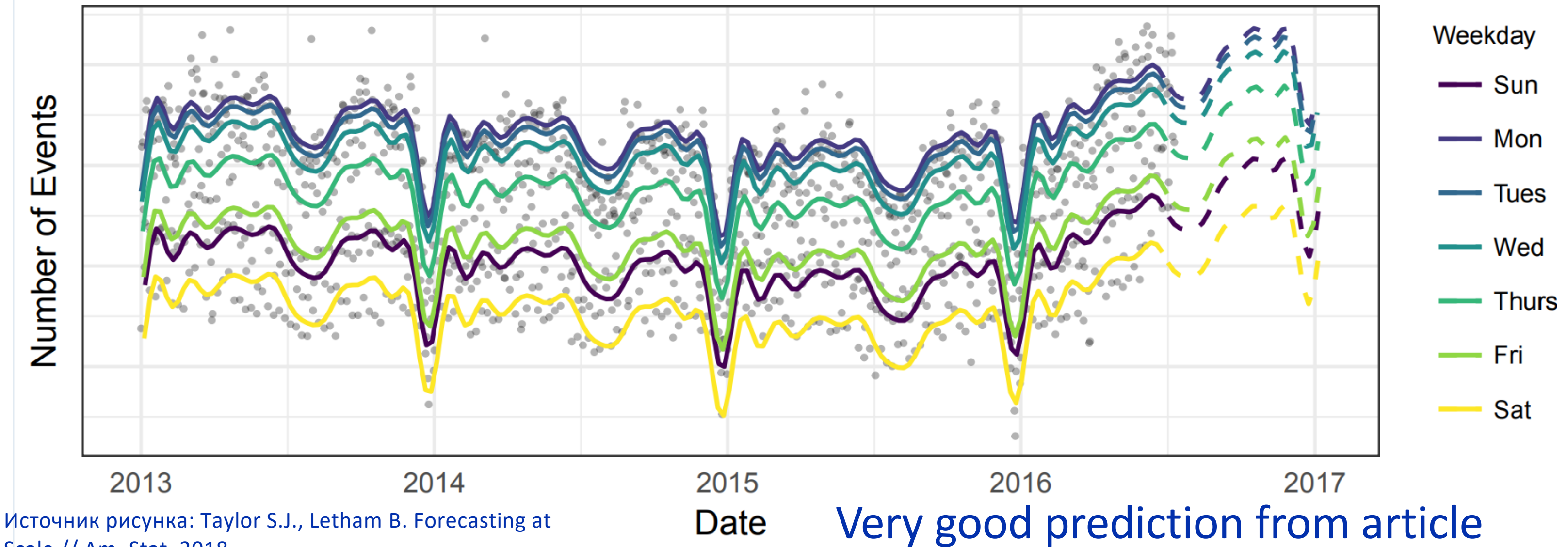
$h(t) \sim N(0, \nu^2)$ + some pre-defined values

Pros:

- Non-stationary timeseries modelling
- Should model (and generate) various timeseries

Cons:

- One scale (daily human behaviour)



Источник рисунка: Taylor S.J., Letham B. Forecasting at Scale // Am. Stat. 2018

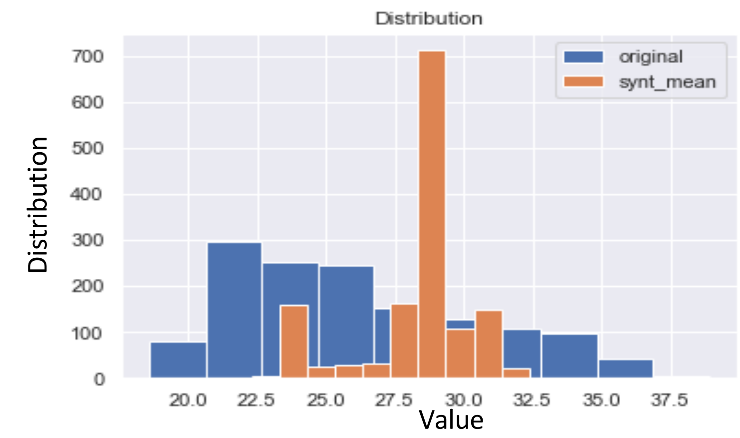
Very good prediction from article

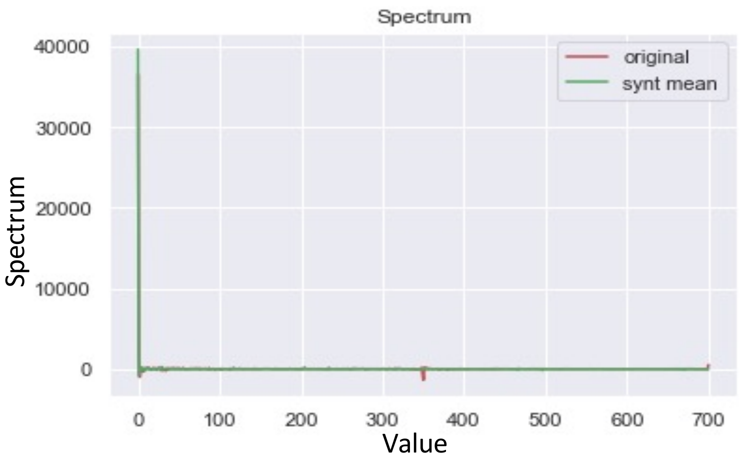
Prophet (I do not want to think way)

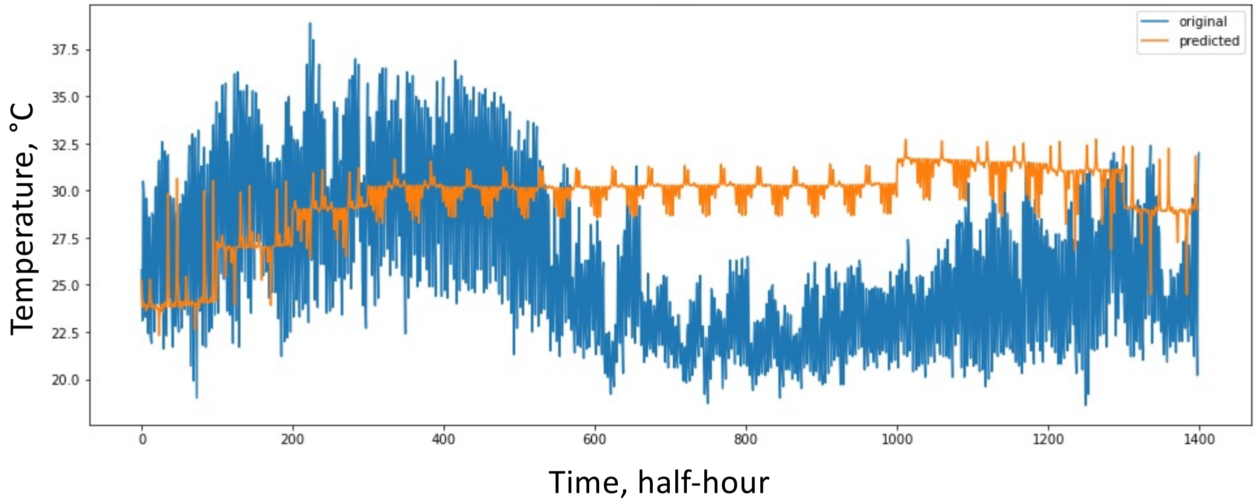
УНИВЕРСИТЕТ ИТМО

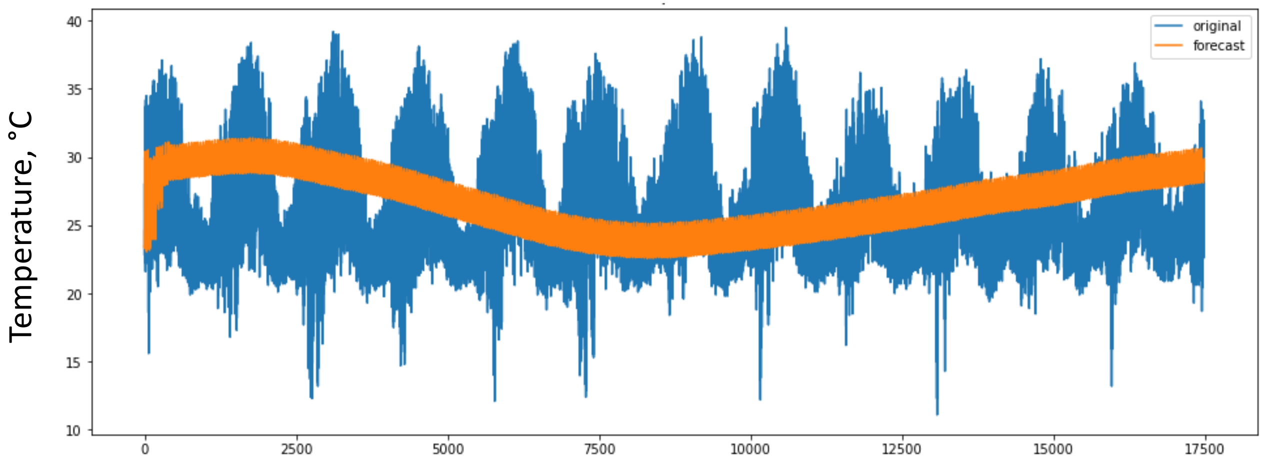
Temperature:

- Nah, not working temperature is not a human









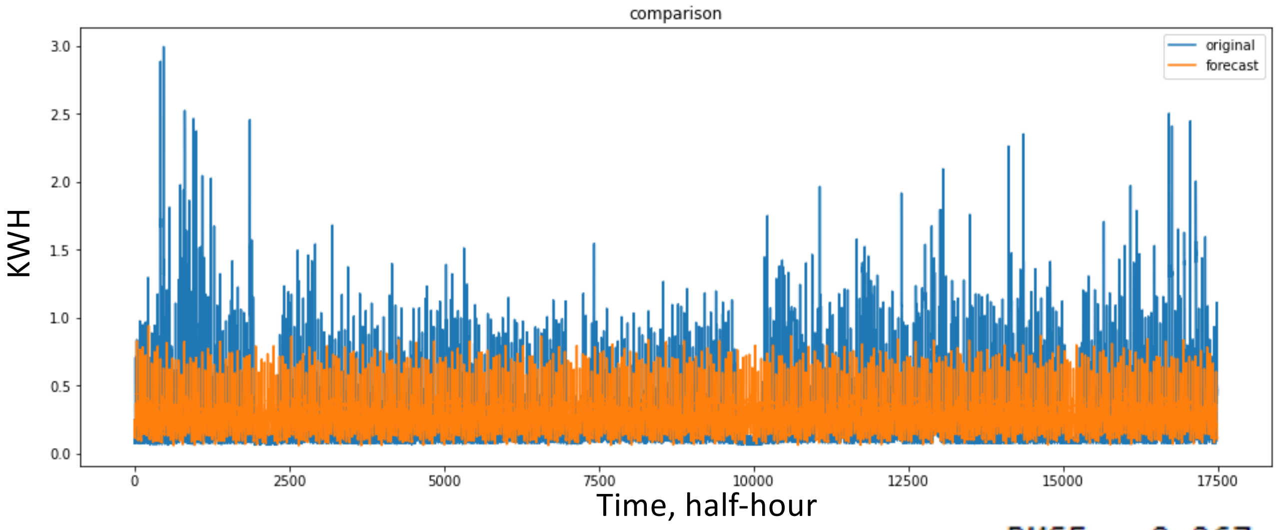
RMSE: 4.839

Prophet (I do not want to think way)

УНИВЕРСИТЕТ ИТМО

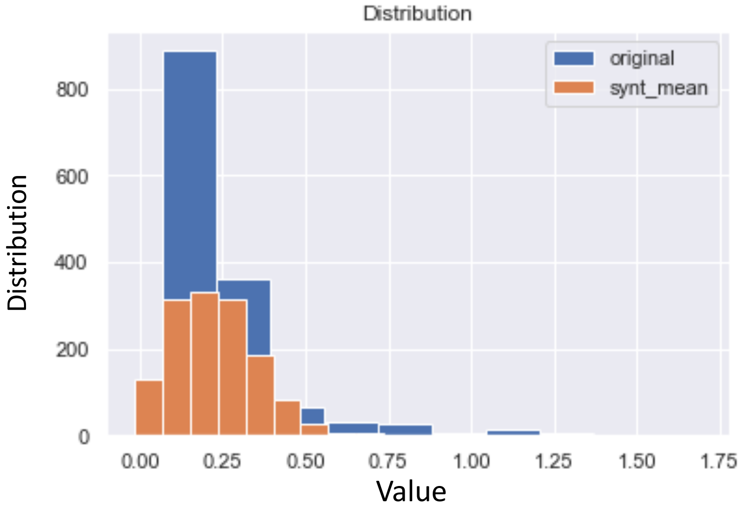
Summary:

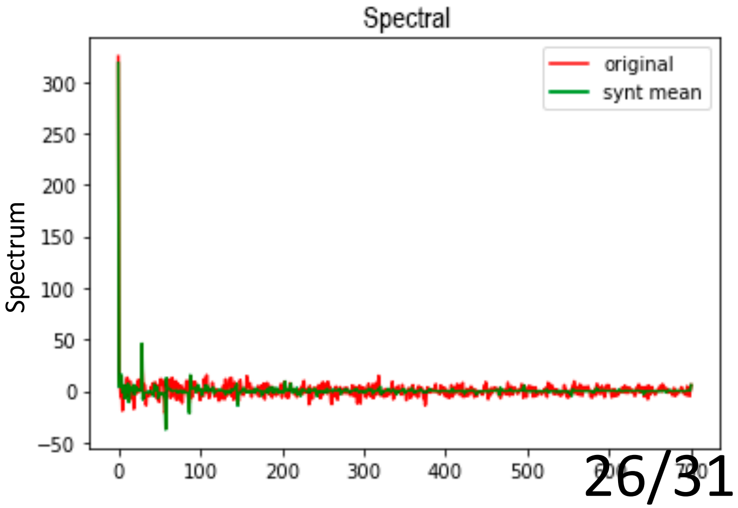
- Prophet cannot be used to model a lot of applications
- Works only for task Facebook designed it
- Holydays (peaks) could not be modelled even if hourly scale is considered



RMSE: 0.267

100 synthetic time series distribution and spectrum





Closed-form expressions

Let the model-expression be defined by a set of tokens:

$$M(\vec{A},t) = \sum_{I}\prod_{J} c_{i,j}(\vec{a}_{i,j},t) \quad I,J\epsilon N$$

$$\vec{A} = (\vec{a}_{i_1,j_1},\ldots) - multi\ parameter$$

Then the **problem** is to find such tokens $c_{i,j}$ with such parameters $\vec{a}_{i,j}$ that the model has the smallest Euclidean distance to the input data P :

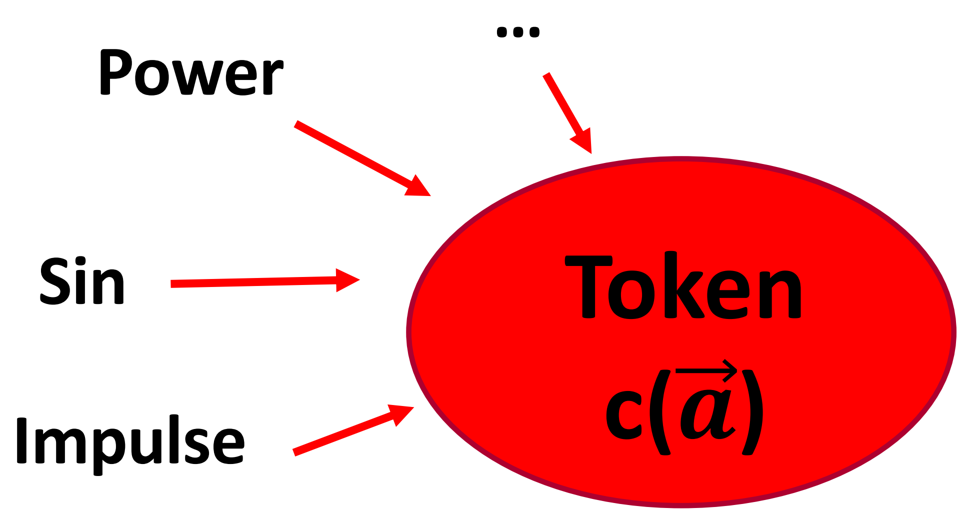
$$argmin_{\vec{A}}\ N(\vec{A}) = \sqrt{\sum_{k=1}^{k=n}\left(P(t_k) - M(\vec{A},t_k)\right)^2}$$

Fitness of the expression

A **token** is a parameterized mathematical (and not very) function and a building block for expressions, for example:

$$c(\vec{a},t) = a_1 Sin(a_2 t + a_3)$$

$$\vec{a} - vector\ of\ parameters$$

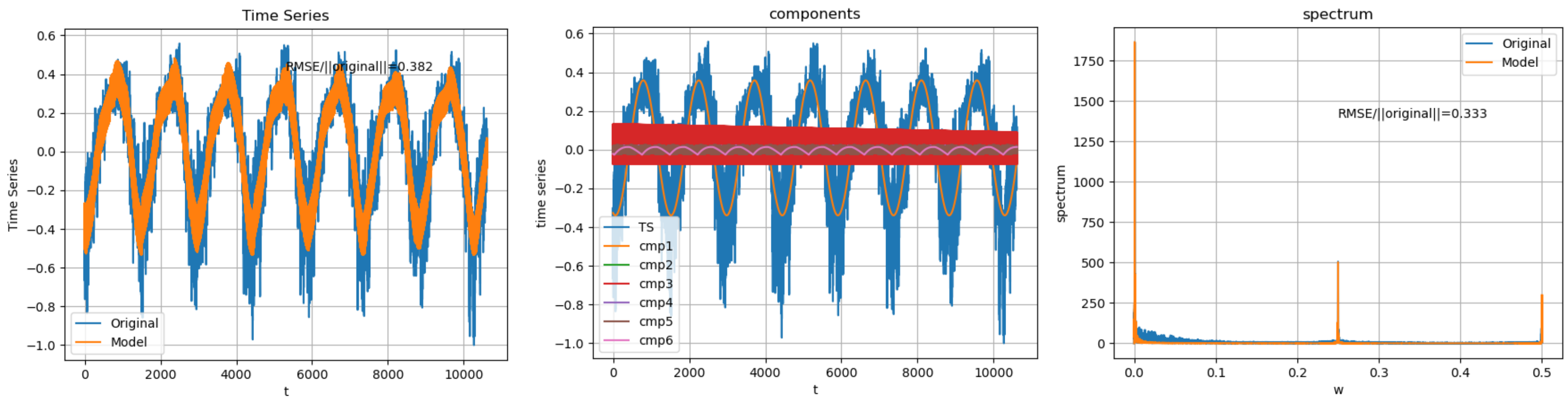


Closed-form expressions

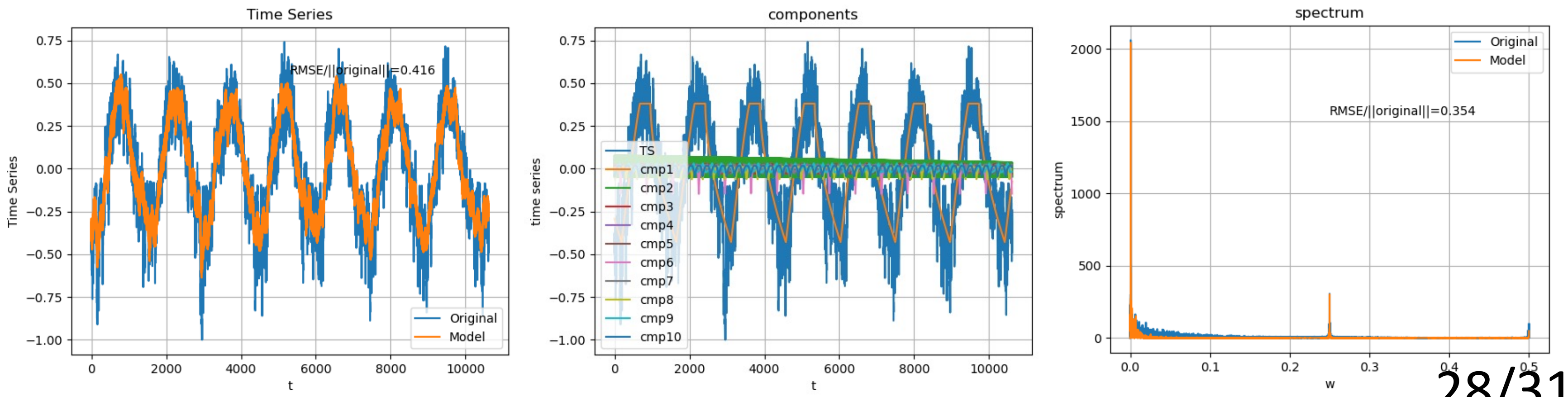
Model:

- We extract trend, seasonal component, pulses and other components using evolutionary algorithm
- Resulting model is a closed-form algebraic expression
- We could model noise for every term as a unique stationary process

Subtropical temperature zone



Moderate temperature zone

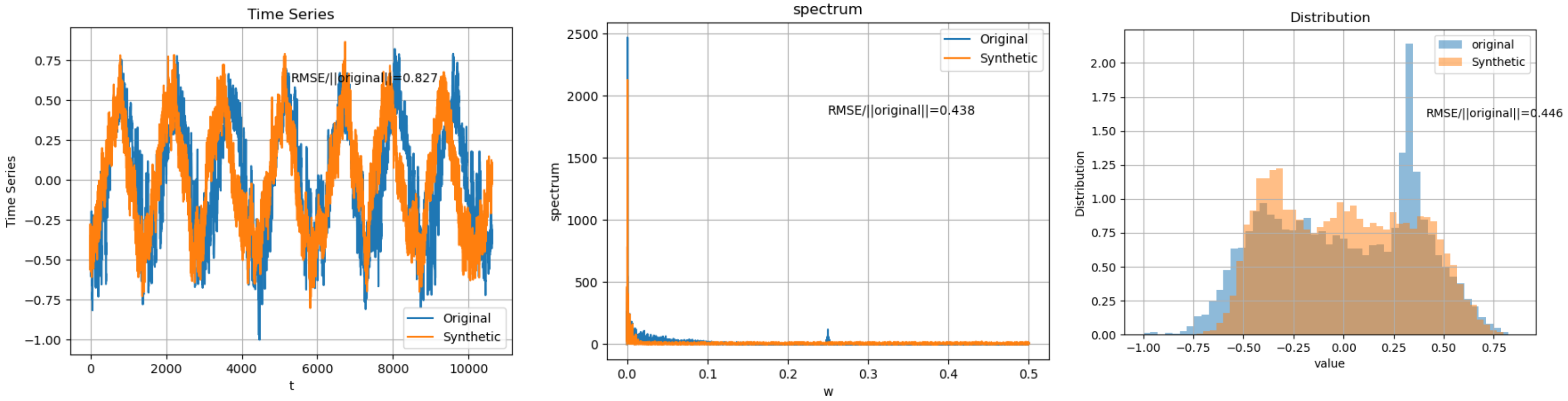


Closed-form expressions

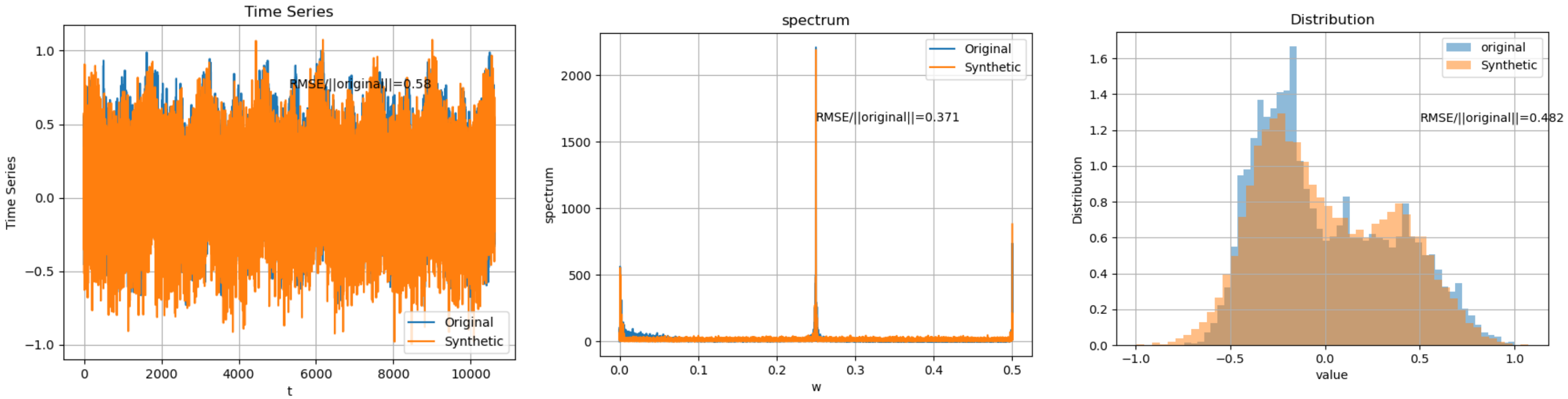
Periodic data:

- You don't have to care about the significance of the parts
- Allows to direct stationary function parameters optimization

Arctic zone temperature



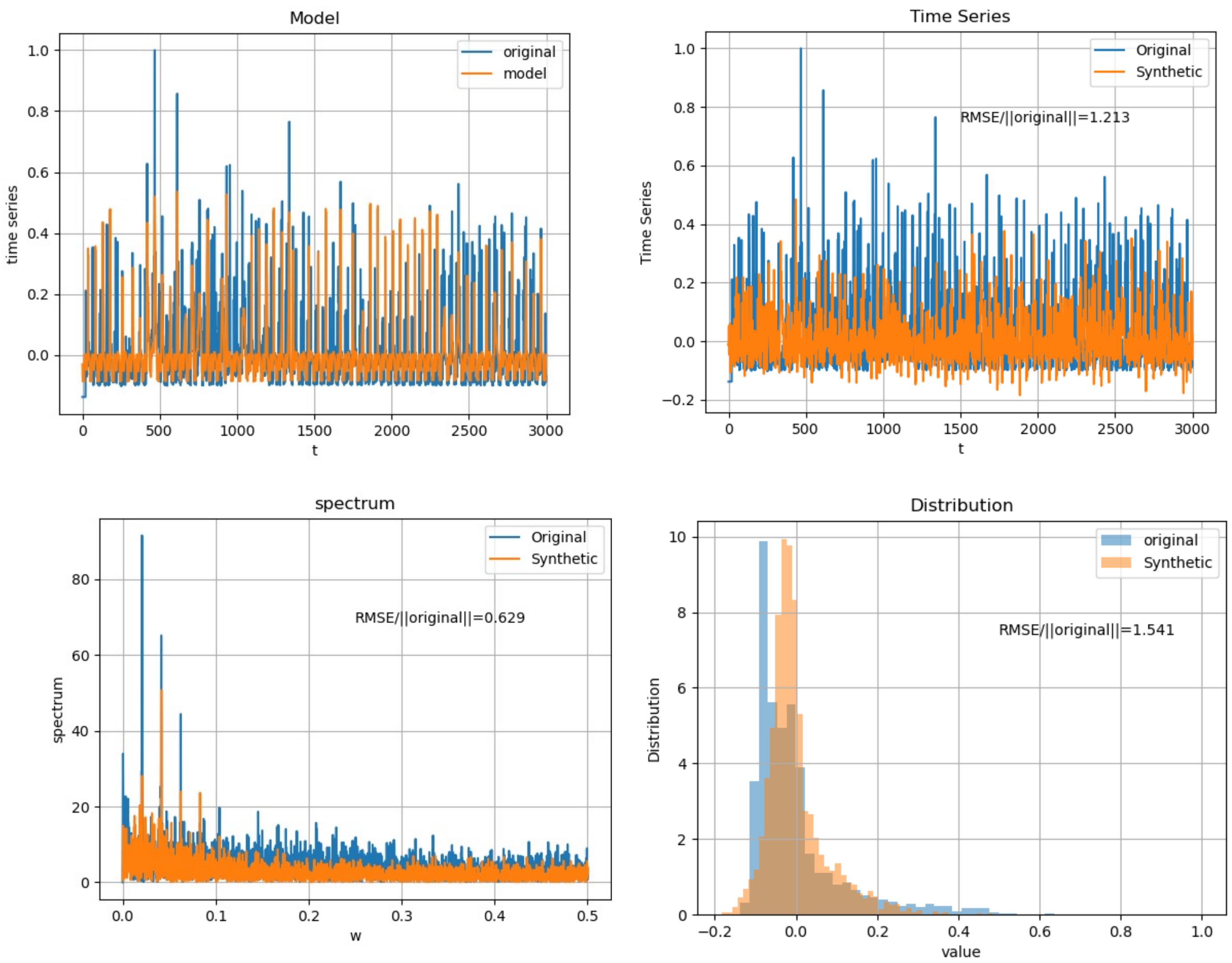
Subequatorial zone temperature



Closed-form expressions

Non-periodic data:

- We use pulses to model different peaks
- Pulse parameters are also may be modelled as the stationary variables
- Since it is required many pulses to model the process, the number of parameters is relatively large and the modelling require fine tuning



Thanks for coming!

www.ifmo.ru

ITsMOre than a UNIVERSITY