# Linear Methods

# for

# Time-Series Analysis

## Motivation

- linear methods
  - can yield complementary, useful information
  - may decide about prerequisites for non-linear methods
  - some are basic ingredients of non-linear methods

- non-linear methods may be overkill
- get acquainted with the pitfalls of data analysis

## Statistical Data Analysis

| ***model-independent*** | ***model-dependent*** |
|---|---|
| • moments of distributions | • model fitting |
| • (in-)equality of distributions | • parameter estimation |
| • correlation | • robust estimation |
| • … | • … |

↓

**descriptive statistics**

## Distribution of Values

**Given**: time series $\boldsymbol{v}$: $v_1, v_2, \ldots, v_N$ of some system observable $\mathbf{x}$

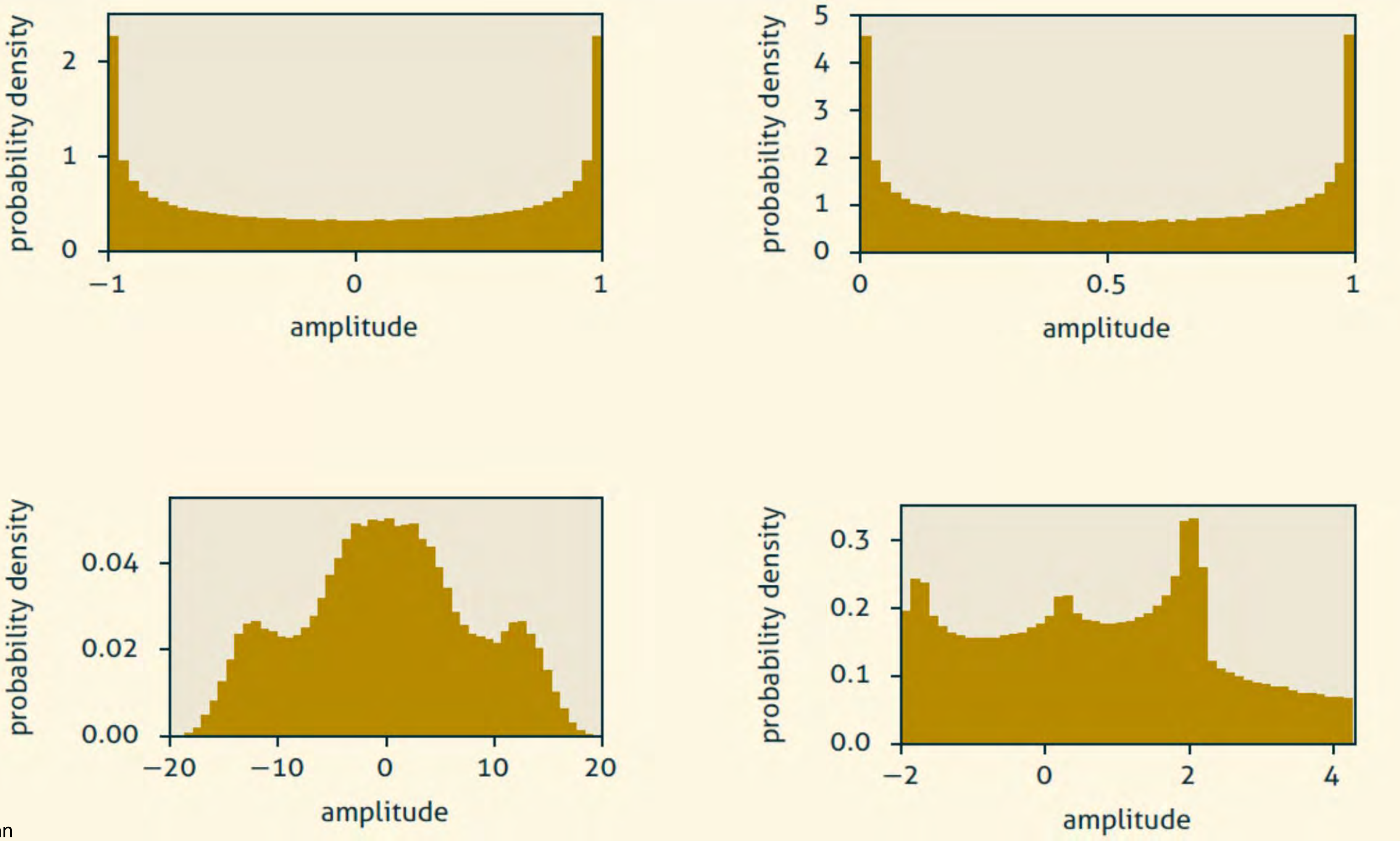
**Assumption:** each value of the time series is independently sampled from some distribution

**Assumption implies:**

- no memory
- no dynamics
- time is not important
- stationarity (definition: later)

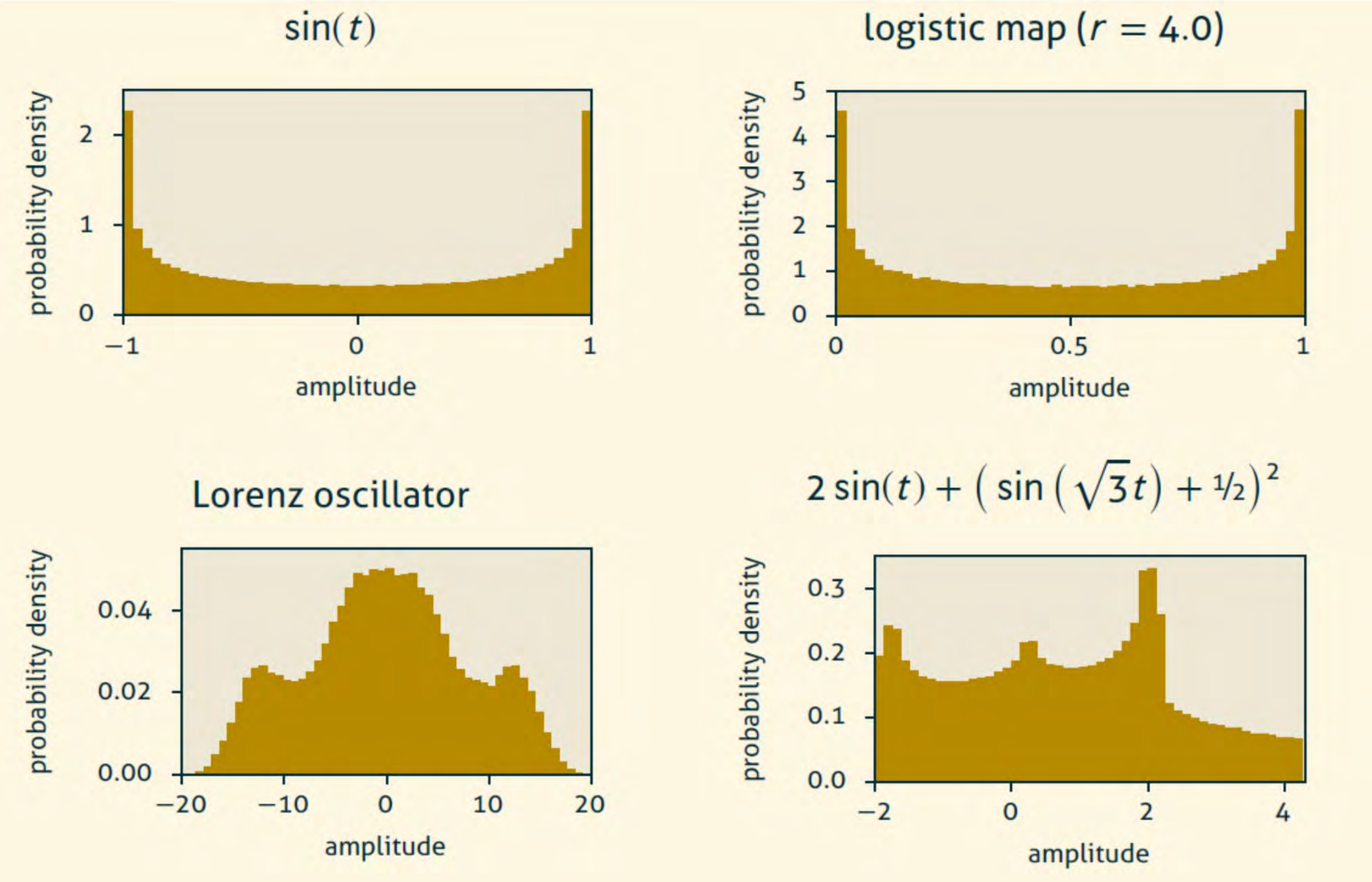
# Distribution of Values

## Examples




from: G. Ansmann

# Distribution of Values **Examples**




from: G. Ansmann

## Statistical Moments of a Distribution

first moment: ***mean***

$$\bar{v} := \frac{1}{N} \sum_{i=1}^{N} v_i$$

*mean vs. expected value:*

- mean $\bar{v}$ is a property of a dataset
- expected value $\langle v \rangle$ is a property of a population
- if a dataset is sampled from some population, $\bar{v}$ is the best estimator for $\langle v \rangle$ (of that population)

(law of large numbers)

## Statistical Moments of a Distribution

second moment: ***variance***

$$\sigma_v^2 := \frac{1}{N-1} \sum_{i=1}^{N} (v_i - \bar{v})^2$$

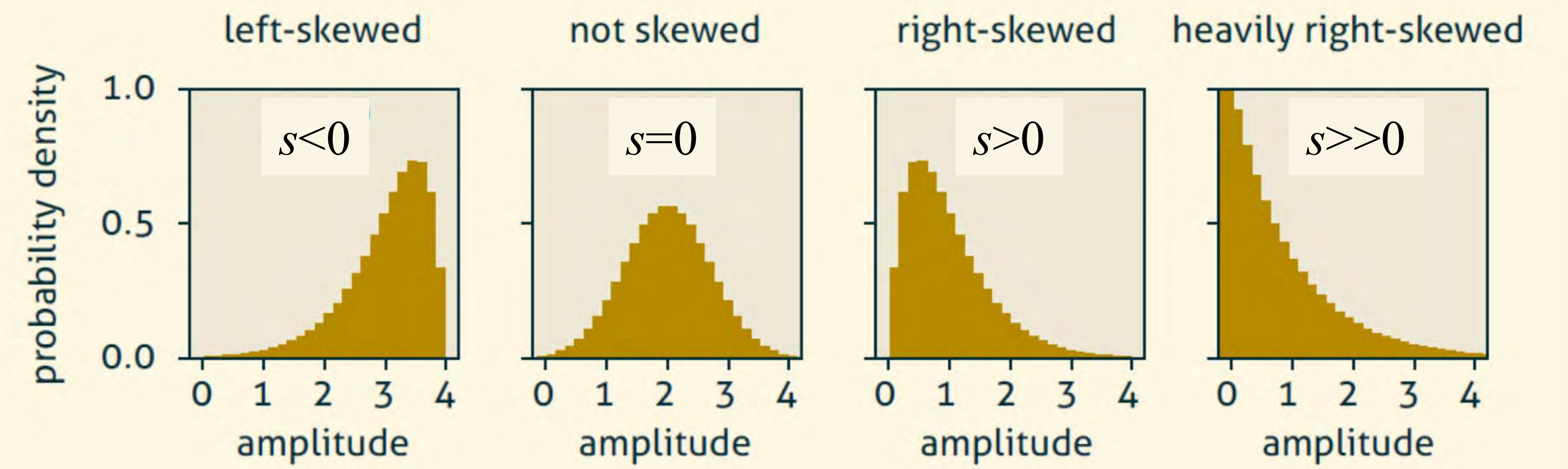
width of the distribution, *variability of the time series*

σ: standard deviation
normalization factor:
$N$-1: estimating variance from a dataset
$N$ : variance of a population

## Statistical Moments of a Distribution

third moment: ***skewness***

$$s_v := \frac{1}{N} \sum_{i=1}^{N} \left( \frac{v_i - \bar{v}}{\sigma_v} \right)^3$$



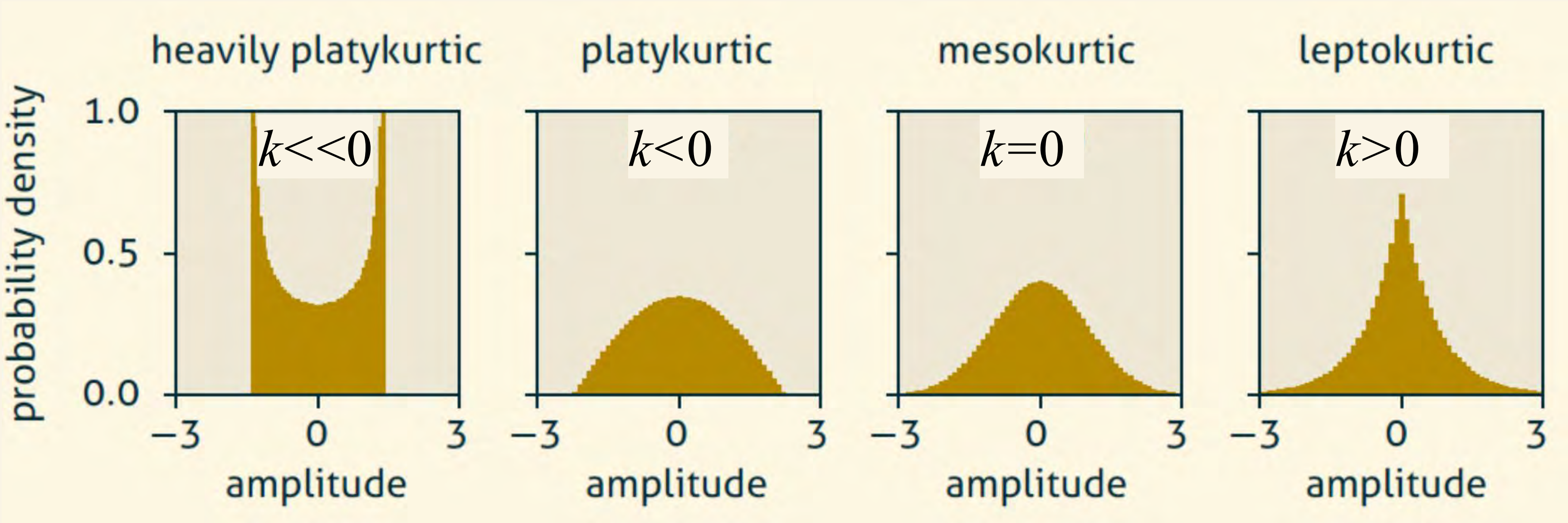

from: G. Ansmann

$s = 0$ for any symmetric distribution

## Statistical Moments of a Distribution

fourth moment: ***kurtosis***

$$k_v := \frac{1}{N}\sum_{i=1}^{N}\left(\frac{v_i - \bar{v}}{\sigma_v}\right)^4 - 3$$




from: G. Ansmann

the normal distribution has $k = 0$

# Statistical Moments of a Distribution

## *interpreting skewness and kurtosis*

- typical noise is a superposition of many small effects
  → typical noise is approximately normally distributed (central limit theorem)

- normal distribution is symmetric and mesokurtic

- significantly non-zero skewness and kurtosis hint at
  - non-linearity of measurement
  - dynamics
  - non-linear dynamics
  - extremes
  - …

## Statistical Tests — Example: skewness

- *assumption / prerequisite:*
  data independently sampled from some population

- *null hypothesis:*
  population not skewed

- *p-value / error probability / significance:*
  probability to find observed skewness
  in a population complying with the null hypothesis
  ↝ probability that null hypothesis is true

*typical procedure:*

1. choose significance threshold $\alpha$, e.g., $\alpha = 0.05$
2. if $p < \alpha$, reject null hypothesis, e.g., consider data skewed

## Statistical Tests

## Example: skewness

***Beware the prerequisites !***

significance values are meaningless if assumptions are not fulfilled

results for skewness test for $\{\sin(t) | t \in T\}$

| $T$ | $p$ |
|---|---|
| (0.00, 0.01, …, 9.00) | $4 \cdot 10^{-9}$ |
| (0.00, 0.01, …, 40.00) | 0.02 |
| (0.00, 0.01, …, 41.00) | 0.002 |
| (0.0, 0.1, …, 9.0) | 0.05 |
| (0, 1, …, 100) | 0.95 |

problem: data not independent !

## Statistical Tests — Comparing Distributions

### *Comparing means* — *Student's t-test*

**Given**: time series $\boldsymbol{v}$: $v_1, v_2, \ldots, v_{N_v}$ and $\boldsymbol{w}$: $w_1, w_2, \ldots, w_{N_w}$
and respective means

$$t = \frac{\bar{v} - \bar{w}}{\sigma_{vw}}$$

where

$$\sigma_{vw} = \sqrt{\frac{\sum_{i=1}^{N_v} (v_i - \bar{v})^2 + \sum_{i=1}^{N_w} (w_i - \bar{w})^2}{N_v + N_w + 2} \left( \frac{1}{N_v} + \frac{1}{N_w} \right)}$$

*p*-value: tables or incomplete beta-function

## Statistical Tests — Comparing Distributions

### *Comparing variances* — *F-test*

**Given**: time series $\boldsymbol{v}$: $v_1, v_2, \ldots, v_{N_v}$ and $\boldsymbol{w}$: $w_1, w_2, \ldots, w_{N_w}$
and respective variances

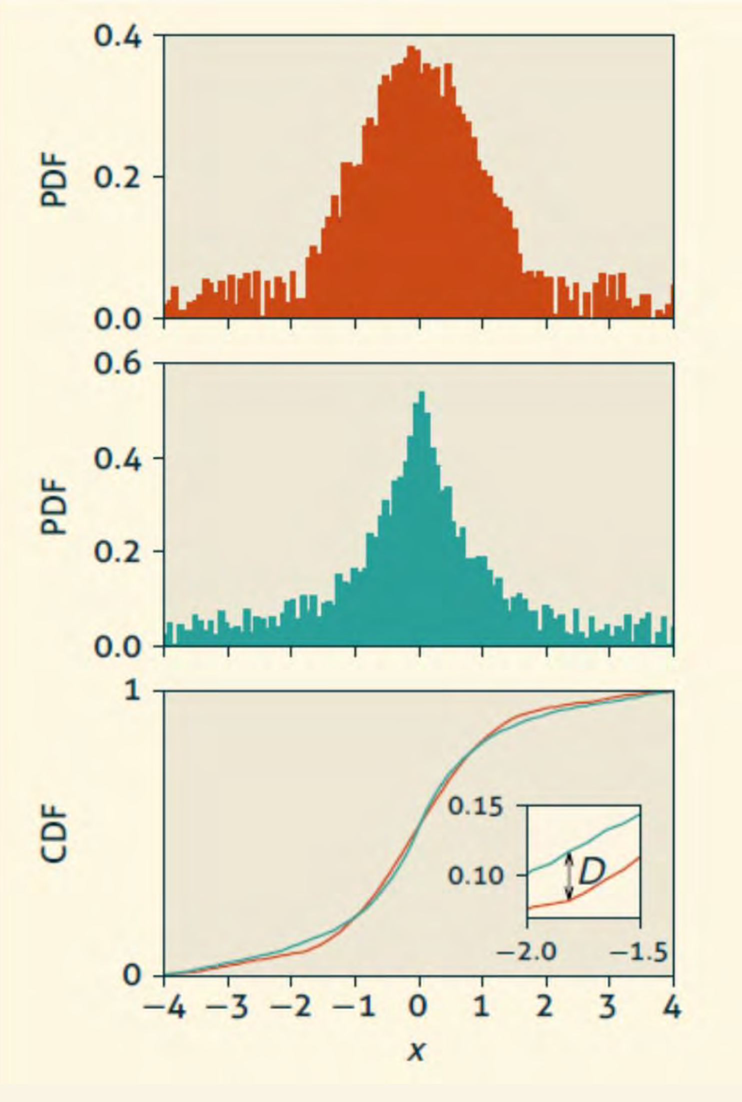
$$F = \frac{\sigma_v^2}{\sigma_w^2}$$

*p*-value: tables or incomplete beta-function

## Statistical Tests

## Comparing Distributions

### *Kolmogorov-Smirnov (KS) test*




from: G. Ansmann

based on cumulative distribution functions:

$$\mathrm{CDF}(x) := \int_{-\infty}^{x} \mathrm{PDF}(\tilde{x}) \mathrm{d}\tilde{x}$$

significance obtained from maximal distance between CDFs

$$D := \max_x |\mathrm{CDF}_1(x) - \mathrm{CDF}_2(x)|$$

*p*-value: tables

## Statistical Tests — Example: KS-test

***Beware the prerequisites (once more) !***

significance values are meaningless if assumptions are not fulfilled

results for comparing $\{\sin(t) | t \in T_1\}$ with $\{\sin(t) | t \in T_2\}$

| $T_1$ | $T_2$ | $p$ |
|---|---|---|
| (0.00, 0.01, …, 9.00) | (3.00, 3.01, …, 12.00) | $6 \cdot 10^{-33}$ |
| (0.00, 0.01, …, 40.00) | (3.00, 3.01, …, 43.00) | $2 \cdot 10^{-5}$ |
| (0.0, 0.1, …, 9.0) | (3.0, 3.1, …, 12.0) | 0.001 |
| (0, 1, …, 100) | (3, 4, …, 103) | 0.99 |

problem: data not independent !

## Statistical Tests Comparing Distributions

### *Pearson's correlation coefficient*

**Given**: time series $\boldsymbol{v}$: $v_1, v_2, ..., v_N$ and $\boldsymbol{w}$: $w_1, w_2, ..., w_N$

covariance

$$\mathrm{cov}_{vw} := \frac{1}{N-1} \sum_{i=1}^{N} \left(v_i - \bar{v}\right)\left(w_i - \bar{w}\right)$$

Pearson's *r*

$$r_{vw} := \frac{\mathrm{cov}_{vw}}{\sigma_v \sigma_w}$$

$r$ = 1: perfect correlation

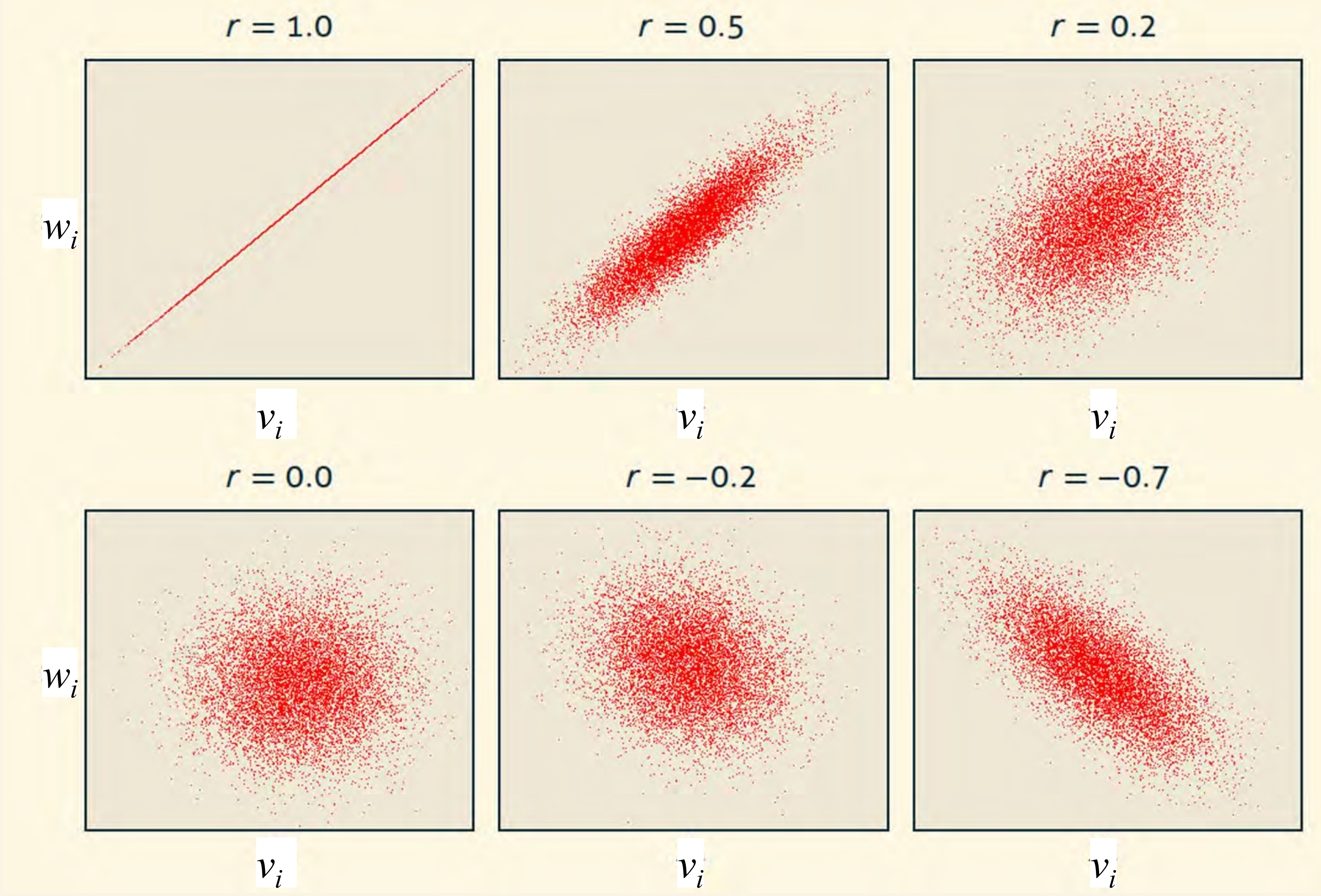
$r$ = 0: no correlation

$r$ = −1: perfect anti-correlation

## Statistical Tests

## Comparing Distributions

### *Pearson's correlation coefficient*




from: G. Ansmann

## Statistical Tests — Cross-Correlation

***extension of Pearson's correlation coefficient***

***Motivation:***

- possible offset in time-dependent data
- sensors may capture dynamics with delay between them

**Given**: time series $\boldsymbol{v}$: $v_1, v_2, \ldots, v_N$ and

shifted time series $\boldsymbol{w}^\tau$: $w_{1+\tau}, w_{2+\tau}, \ldots, w_N$

Cross-correlation (with appropriately truncated time series):

$$C_{vw^\tau} = r_{vw^\tau}$$

symmetry: $$C_{vw^\tau} = C_{wv^{-\tau}}$$

## Statistical Tests Cross-Correlation

### *Intermezzo: application of cross-correlation*

***task:*** find delay and synchrony between two time series

1. find delay that maximizes cross-correlation:

$$\hat{\tau} = \mathrm{argmax}_{\tau} C_{vw^{\tau}}$$

2. use maximized cross-correlation as measure for synchrony

restrictions:

assumes comparable dynamics
assumes "simple" form of synchronization (details later)

## Statistical Tests — Auto-Correlation

Auto-correlation (with appropriately truncated time series):

$$R_{v^\tau} := C_{vv^\tau} = r_{vv^\tau}$$

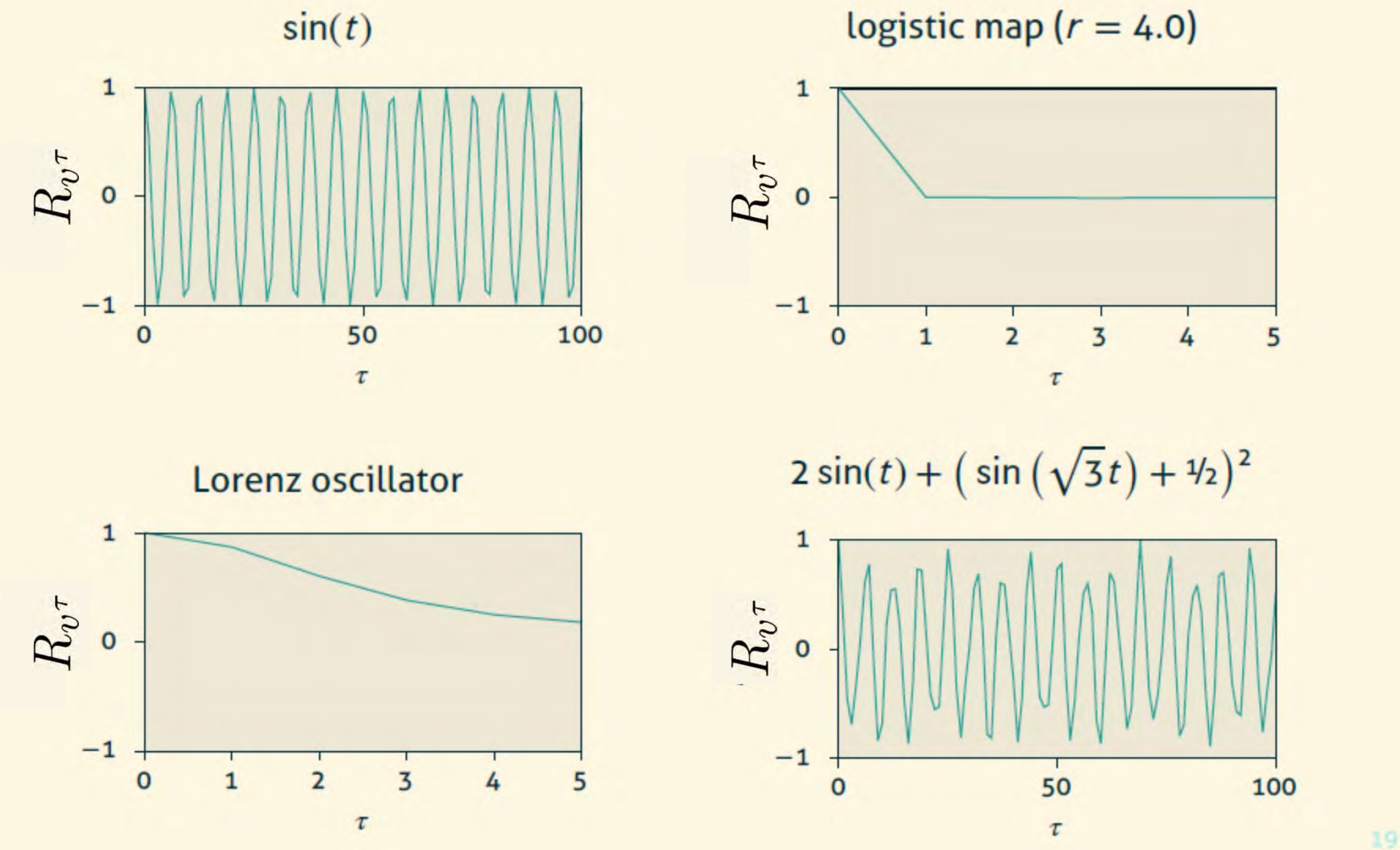
properties:

$$R_{v^\tau} = R_{v^{-\tau}}$$

$$R_{v^{\tau=0}} = 1$$

positive autocorrelation implies some repeating structure in the data

## Statistical Tests Auto-Correlation: Examples




from: G. Ansmann

## Statistical Tests — Rank-based Methods

It is sometimes more appropriate to consider how values rank instead of considering the actual values:

pros: robust against outliers, often fewer constraints on data
cons: information is discarded

| amplitude-based method | rank-based analogous method |
| --- | --- |
| mean | median |
| Pearson's *r* | Kendall's tau, Spearman's rho |
| Kolmogorov-Smirnov test | Mann-Whitney test |

## Stationarity

- Stationarity is a system property!
- definition for time series analysis:

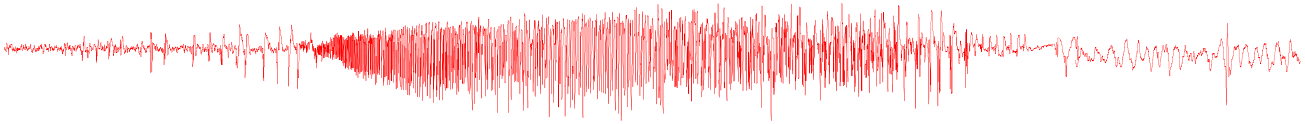
  “*a (stochastic) process is called stationary if the distribution of its states over an ensemble of realizations of that process does not depend on time*”

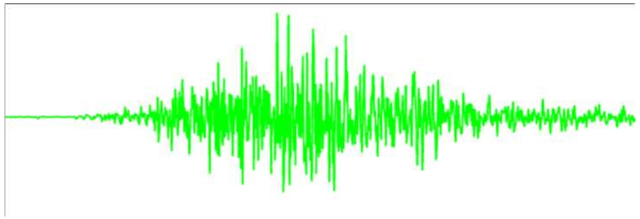
- this implies:

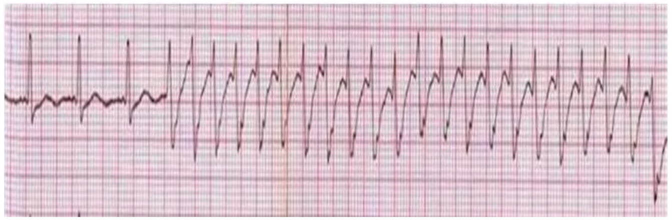
  *constancy of all statistical moments (mean, variance, …) and all joint statistical moments (covariance, …)*

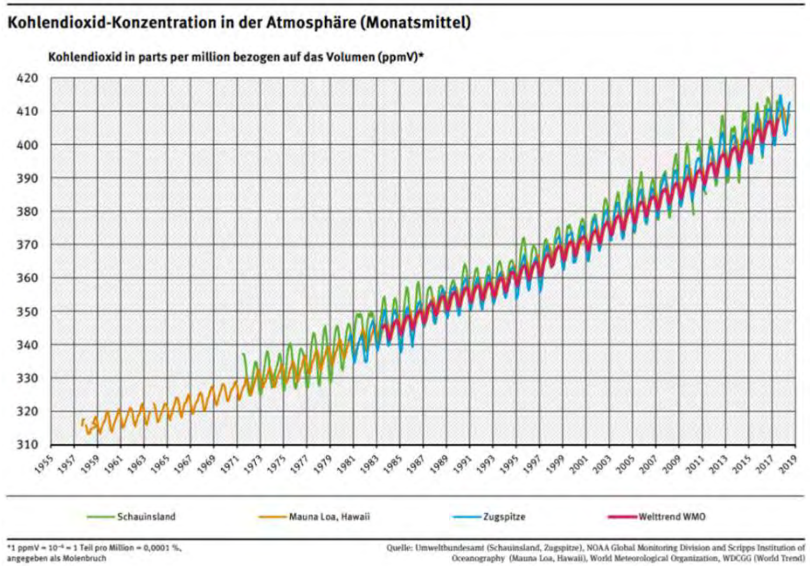
- examples for non-stationary processes:
  - dynamics with changing parameters
  - driven dynamics
  - transient dynamics

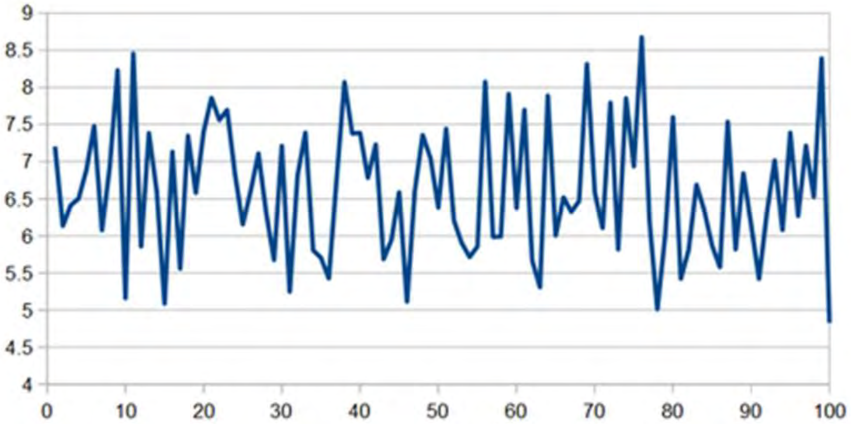
# Stationary or not?

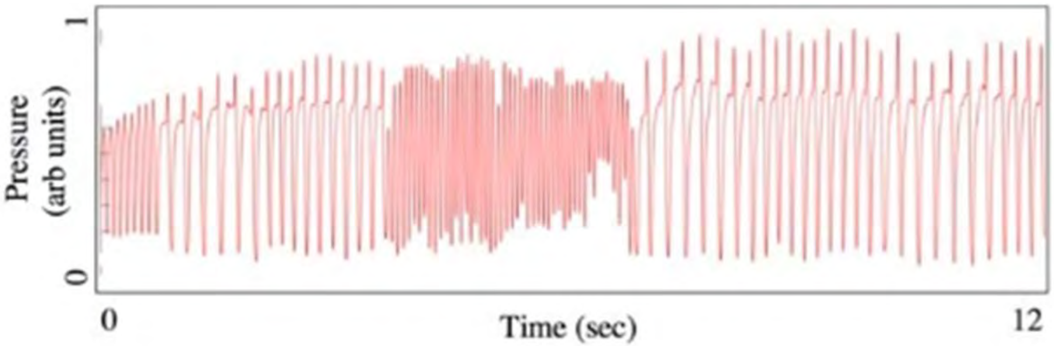


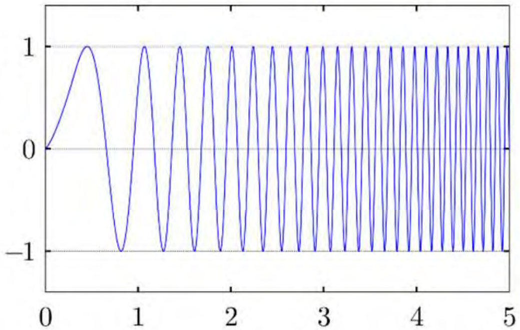








## Stationarity

**prerequisite of most analysis techniques**

- ensures reproducibility of experiments
- required for ergodicity
  (time average ↔ phase space average)
- depends on the time scale:

  on short time scales, an non-stationary process can be approximated as stationary

  on long time scales, instationarities may be regarded as parts of the dynamics or a driver

# Stationarity

## strong stationarity

*“a (stochastic) process is called **strongly stationary** if the distribution of its states over an ensemble of realizations of that process does not depend on time”*

## weak stationarity

*“a (stochastic) process is called **weakly stationary** if its mean, variance, and covariances do not depend on time”*

## Frequency Spectrum

## Identifying hidden periodicities

**Assumption:**

The time series can be decomposed into periodic components

**This implies**

- periodicity, quasiperiodicity
- no chaos
- memory

## Frequency Spectrum

## Fourier transform

### Continuous Fourier transform:

$$\hat{v}(\omega) := \int_{-\infty}^{\infty} v(t) \exp\left(-i\omega t\right) \mathrm{d}t; \quad \omega = 2\pi f$$

### Discrete Fourier transform:

$$\hat{v}_k := \sum_{t=0}^{N} v_t \exp\left(\frac{-ikt}{N}\right)$$

Numerical realization:
- Fast Fourier Transform (**FFT**)
- beware how the output is aligned

# Properties of the Fourier transform

**Convolution theorem** $\widehat{v * w} := \hat{v} \cdot \hat{w}$

**Correlation theorem** $C_{vw} := \hat{v}^* \cdot \hat{w}$

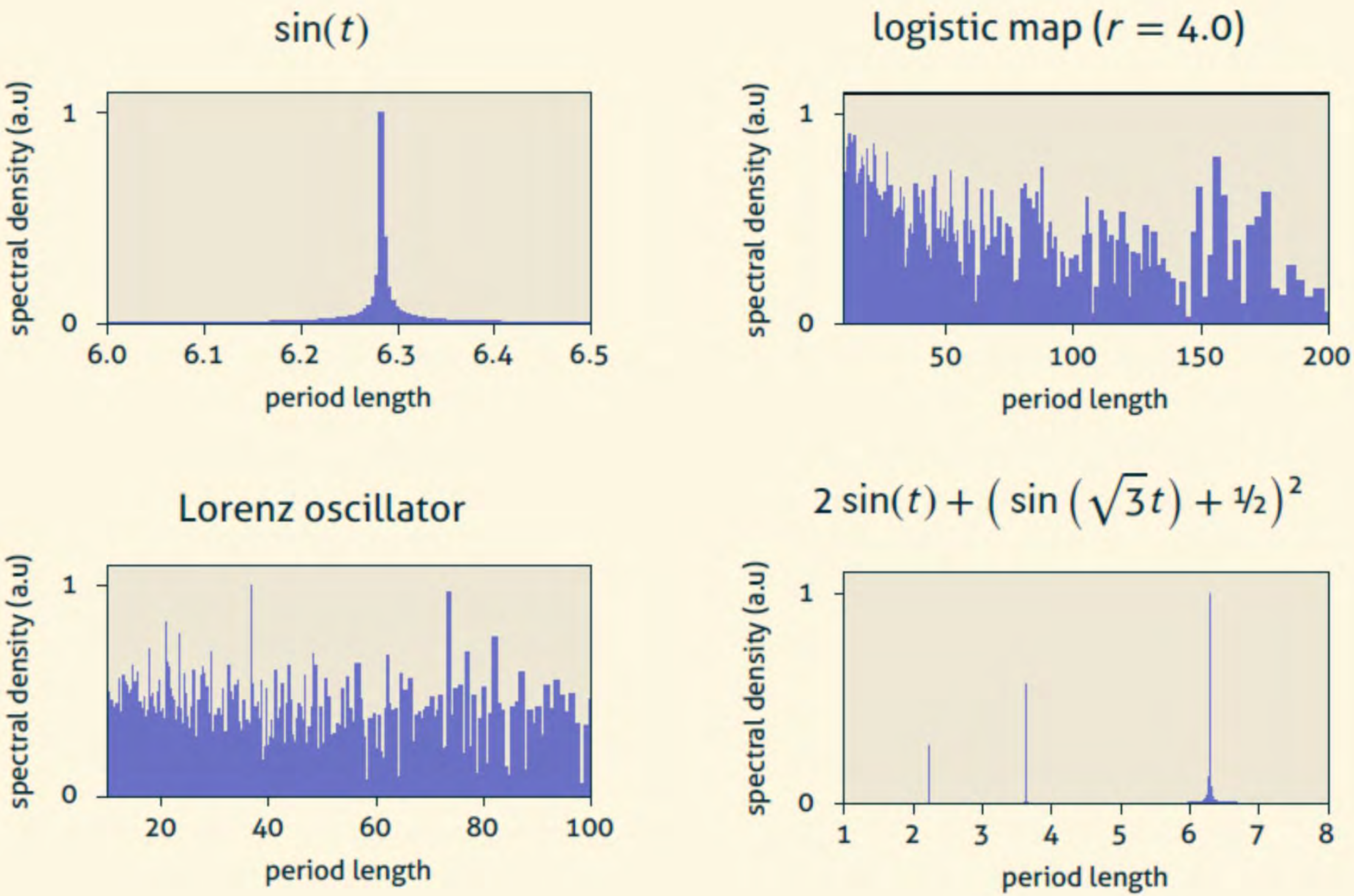
**Wiener-Khinchin theorem** $\widehat{R_v} = \widehat{C_{vv}} = \hat{v}^* \cdot \hat{v} = |v|^2$

**Plancharel theorem** $\sum_{t=1}^{N} v_t^* \cdot w_t \propto \sum_{k=1}^{N} \hat{v_k}^* \cdot \hat{w}_k$

**Parseval's theorem** $\sum_{t=1}^{N} |v_t|^2 \propto \sum_{k=1}^{N} |\hat{v_k}|^2$

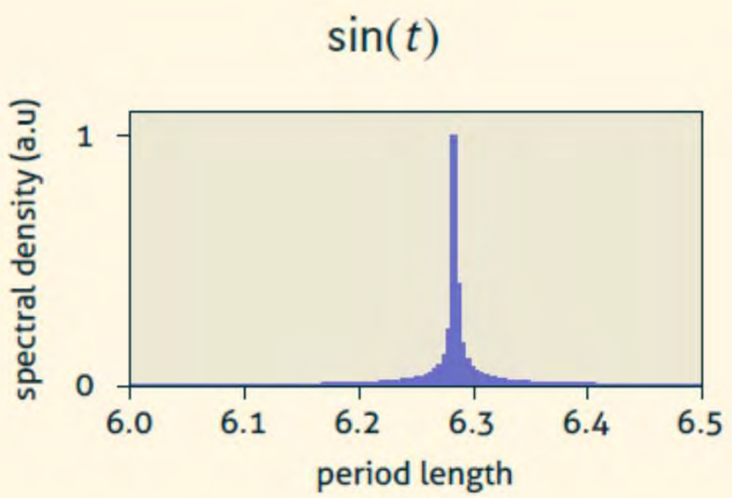
… and respective analogues for the inverse Fourier transform

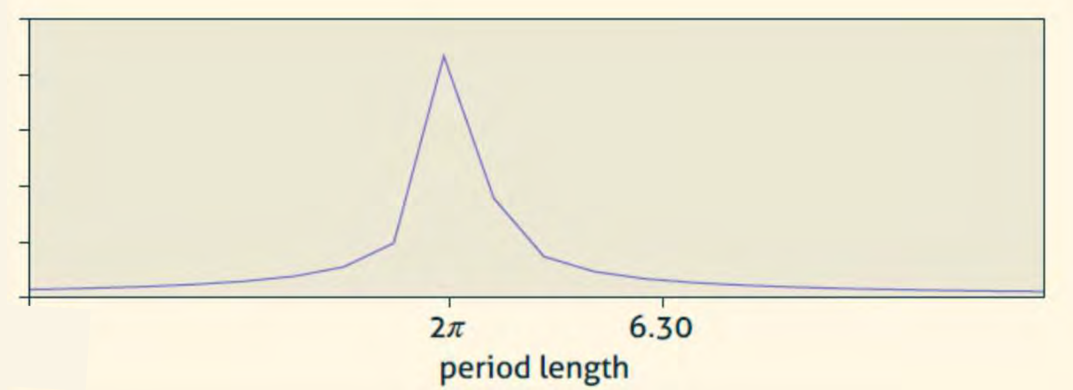
# Fourier transform Examples




from: G. Ansmann

## Fourier transform Spectral Leakage



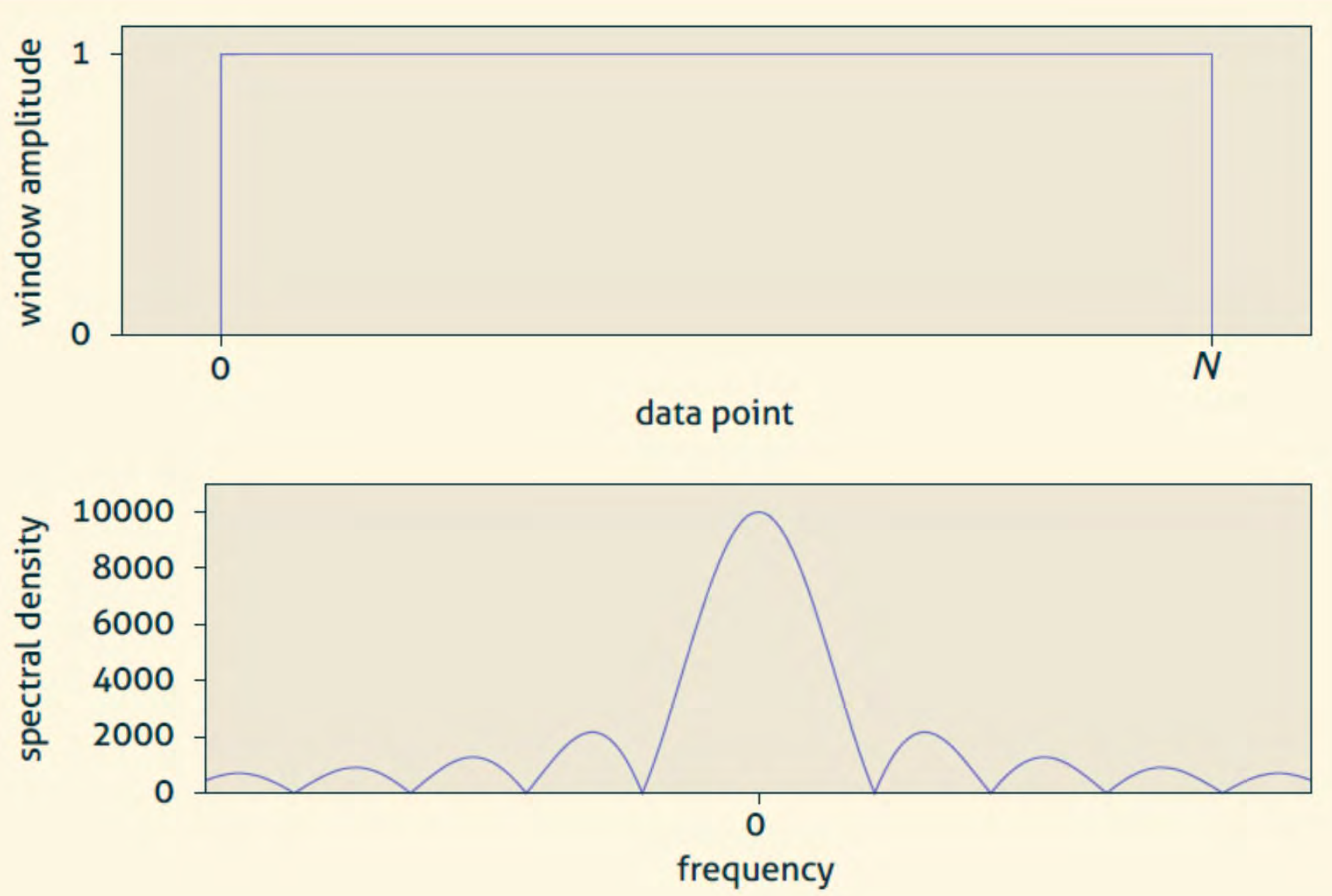




from: G. Ansmann

problem:

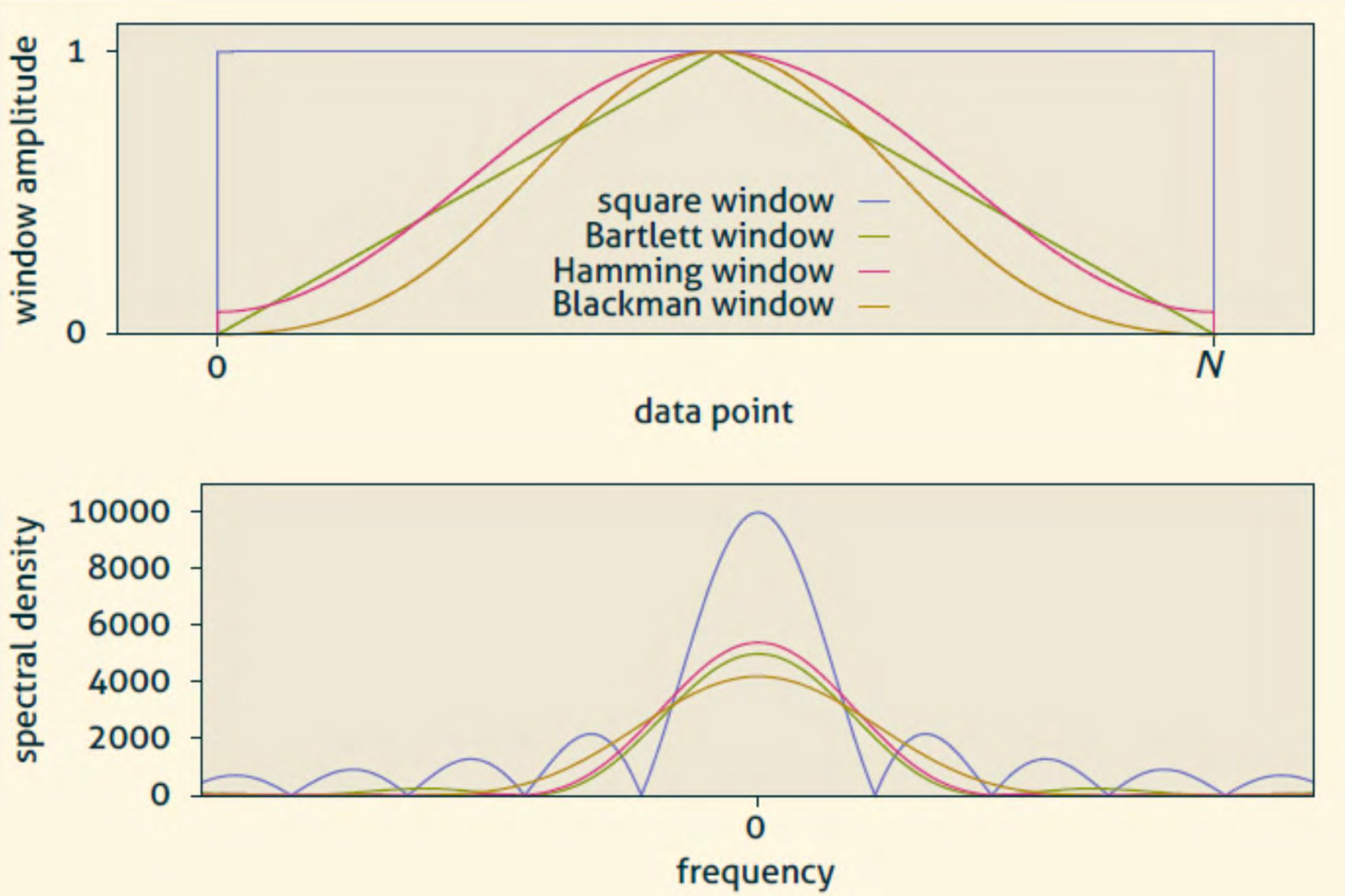
- integration limits (- ∞ to + ∞) ignored
- effectively: convolution of an infinitely long periodic signal with a rectangular window of finite ($N$) size → spectral leakage

# Fourier transform Spectral Leakage and Windowing



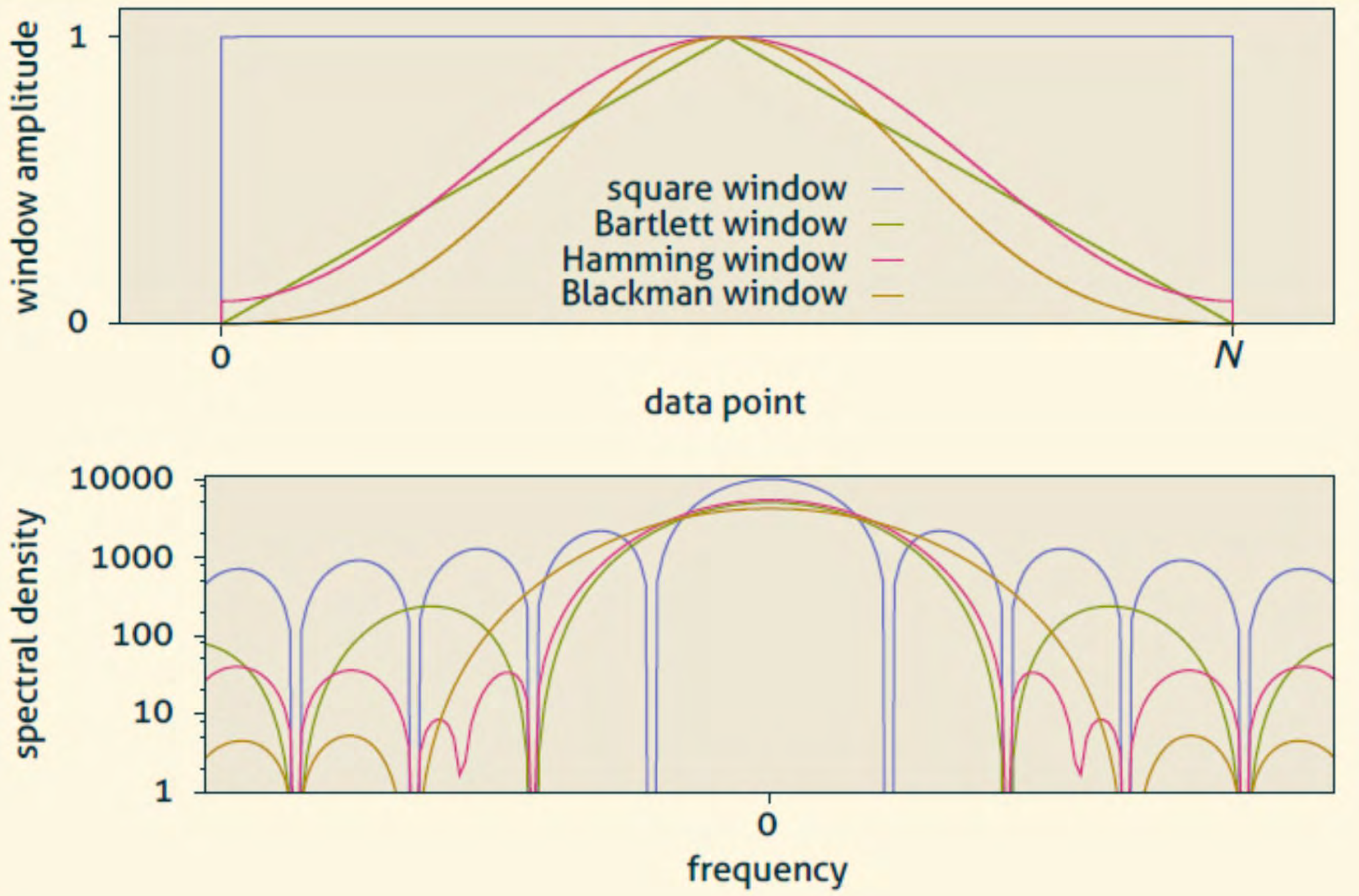

from: G. Ansmann

## Fourier transform Spectral Leakage and Windowing



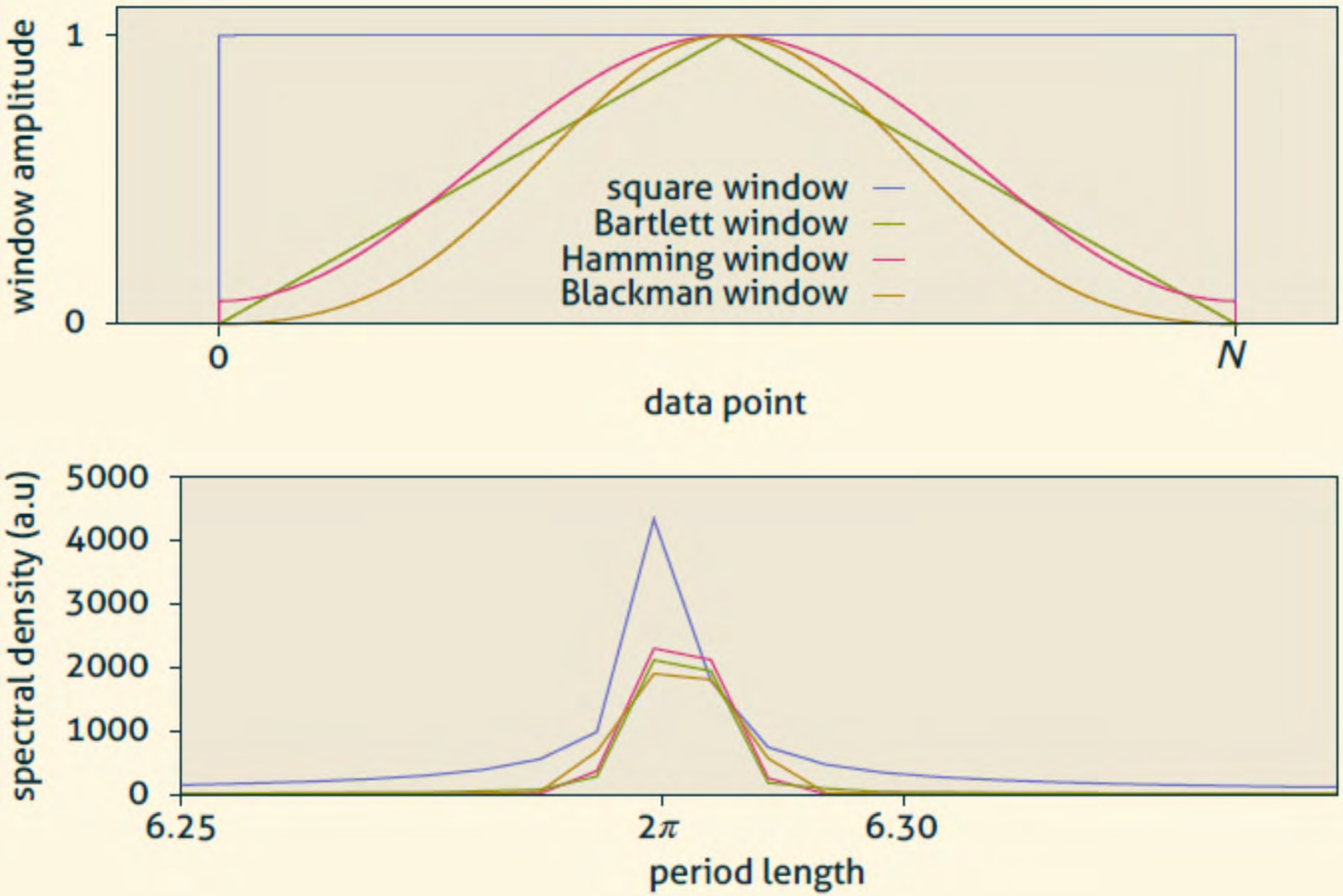

from: G. Ansmann

# Fourier transform Spectral Leakage and Windowing




from: G. Ansmann

# Fourier transform Spectral Leakage and Windowing




from: G. Ansmann

## Fourier transform — Uncertainty

**The standard deviation of each Fourier coefficient is as large as its actual value!**

Minimization of uncertainty (ergodicity assumed)

$\longrightarrow$ averaging over moving windows in the time domain

$\longrightarrow$ moving average in the frequency domain

## Linear Stochastic Processes

processes whose realizations depend on chance

- demonstrate limits of linear methods
- contain most real linear processes as a special case
- null model / null hypothesis
- used for data-driven modelling and forecasting

## Linear Stochastic Processes — White Noise

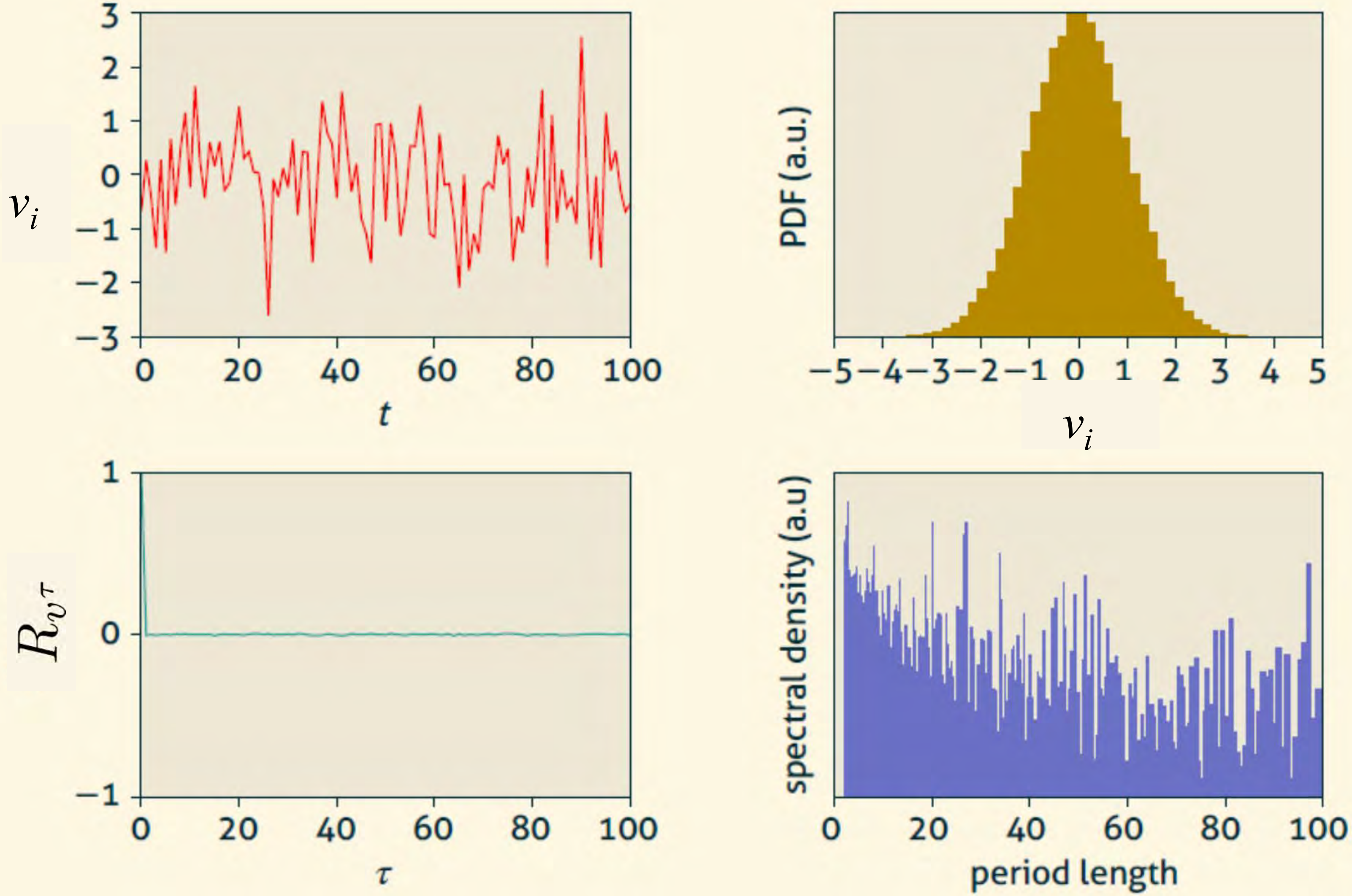
Each sample/value is independently drawn from the same distribution:

$$v_i = \epsilon_i; \quad i = 1, \ldots, N$$

- all frequencies are equally present (analogy: white light)
- autocorrelation is zero, except for a delay of 1
- most often: Gaussian white noise
- basis for the following models.

# Linear Stochastic Processes

## White Noise




from: G. Ansmann

## Linear Stochastic Processes — AR(*k*)-processes

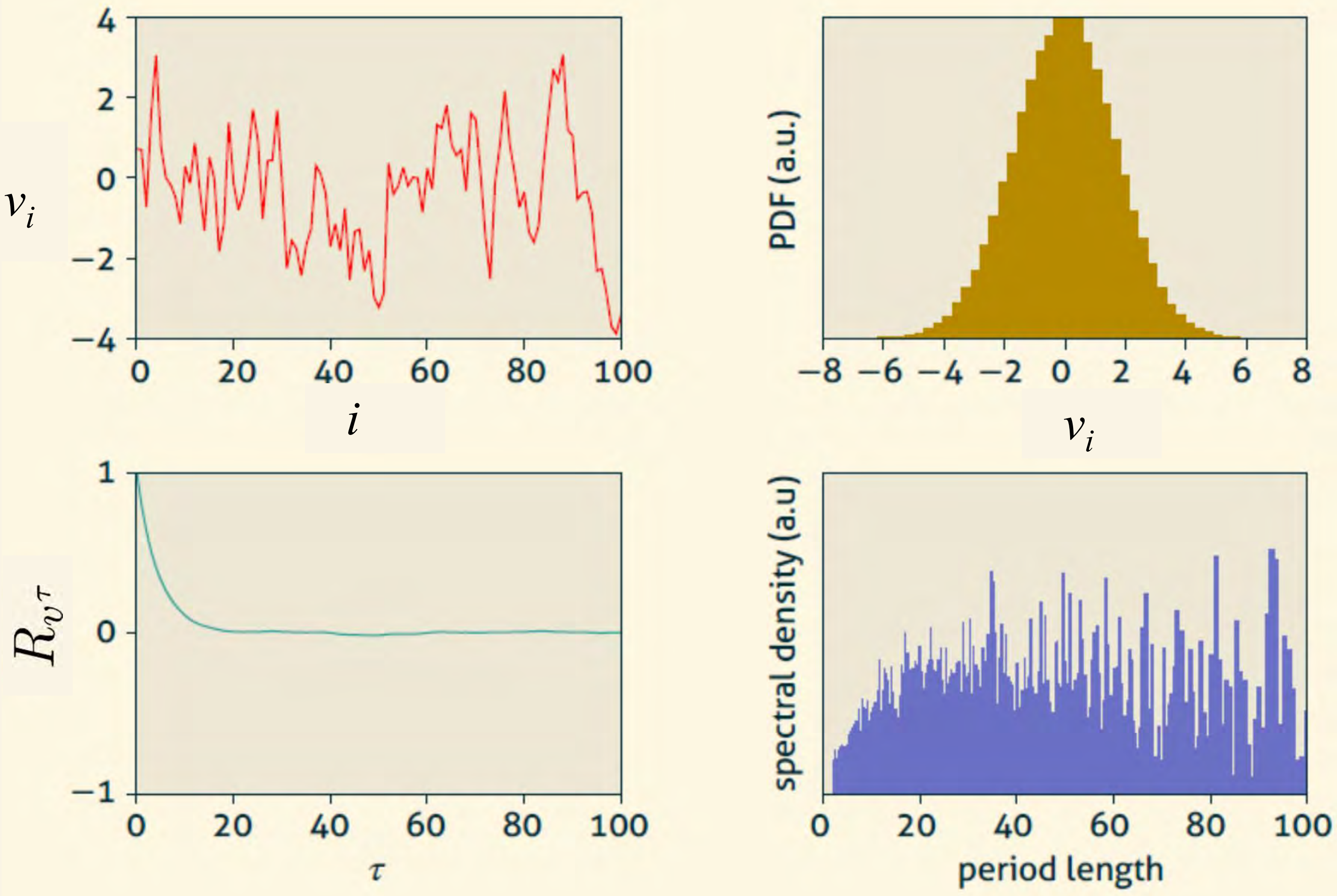
Autoregressive process of order *k*=1 (AR(1))

$$v_i = \alpha v_{i-1} + \epsilon_i; \quad i = 1, \ldots, N$$

***Idea***: Random process with some memory

for $\alpha > 0$, autocorrelation decays exponentially
for $\alpha < 0$, exponentially damped oscillation

# Linear Stochastic Processes — AR(*k*)-processes

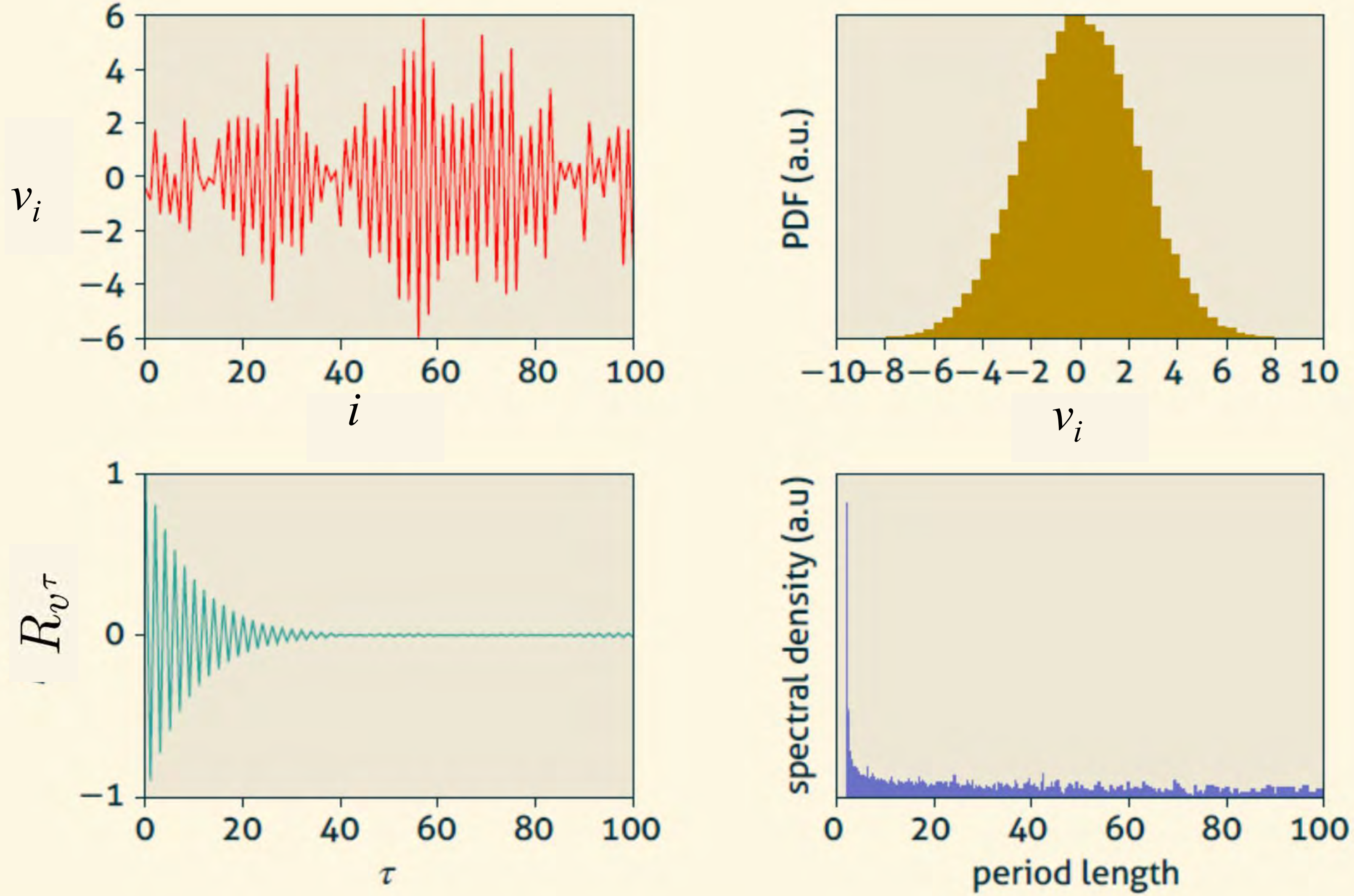



$k = 1$

$\alpha = 0.8$

from: G. Ansmann

# Linear Stochastic Processes AR(*k*)-processes




$k = 1$

$\alpha = -0.9$

from: G. Ansmann

## Linear Stochastic Processes — AR(*k*)-processes

Autoregressive process of order *k* (AR(*k*))

$$v_i = \sum_{j=1}^{k} \alpha_j v_{i-j} + \epsilon_i; \quad i = 1, \ldots, N$$

***Idea***: Random process with some memory

Autocorrelation is superposition of exponential decays
and exponentially damped oscillations

## Linear Stochastic Processes — ARMA(*k,l*)-processes

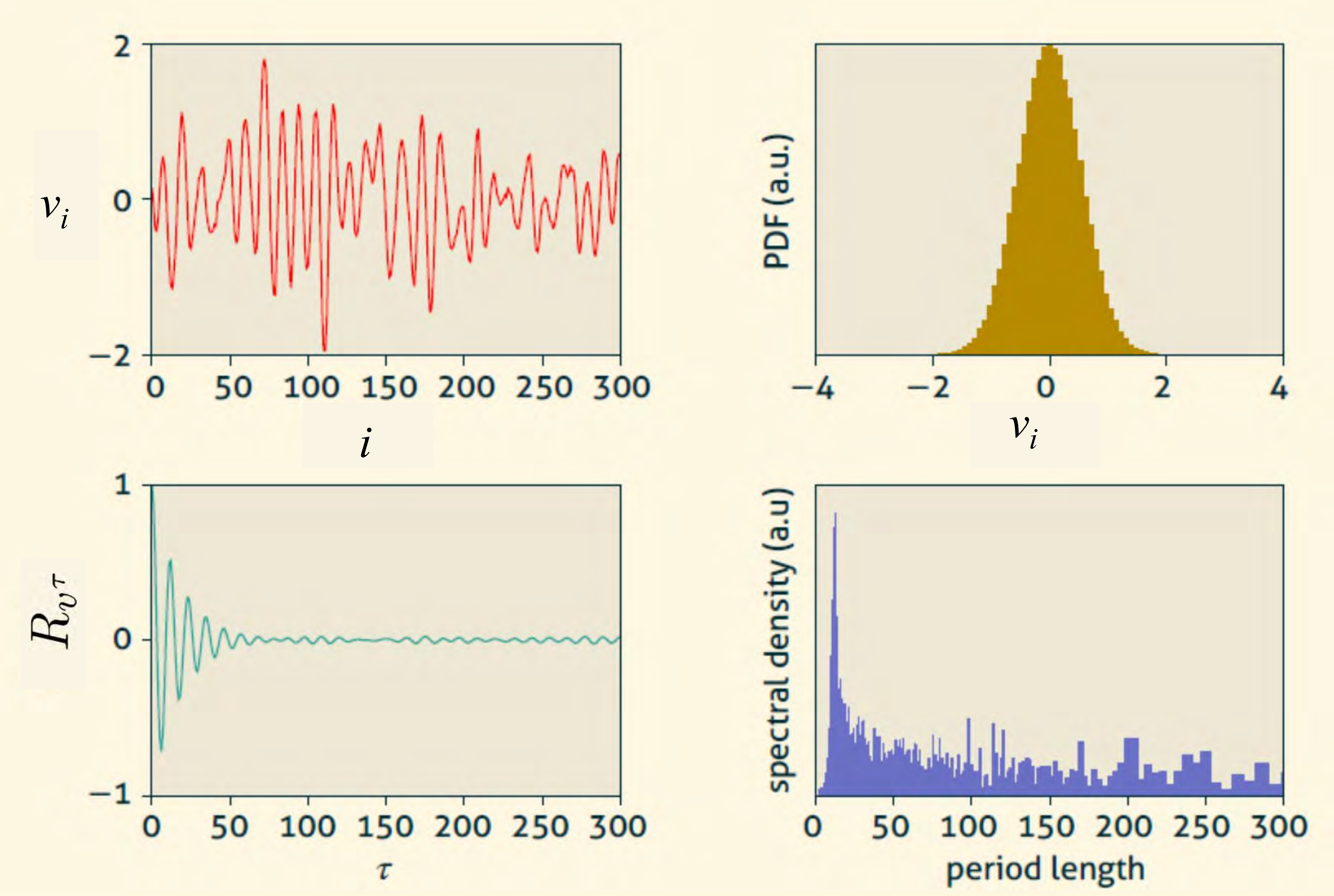
Autoregressive moving-average process of orders *k,l* (AR(*k,l*))

$$v_i = \sum_{j=1}^{k} \alpha_j v_{i-j} + \sum_{m=1}^{l} \beta_m \epsilon_{i-m}$$

***Idea***: Random process with some memory and smoothed noise

# Linear Stochastic Processes ARMA-processes

$k=l=4$



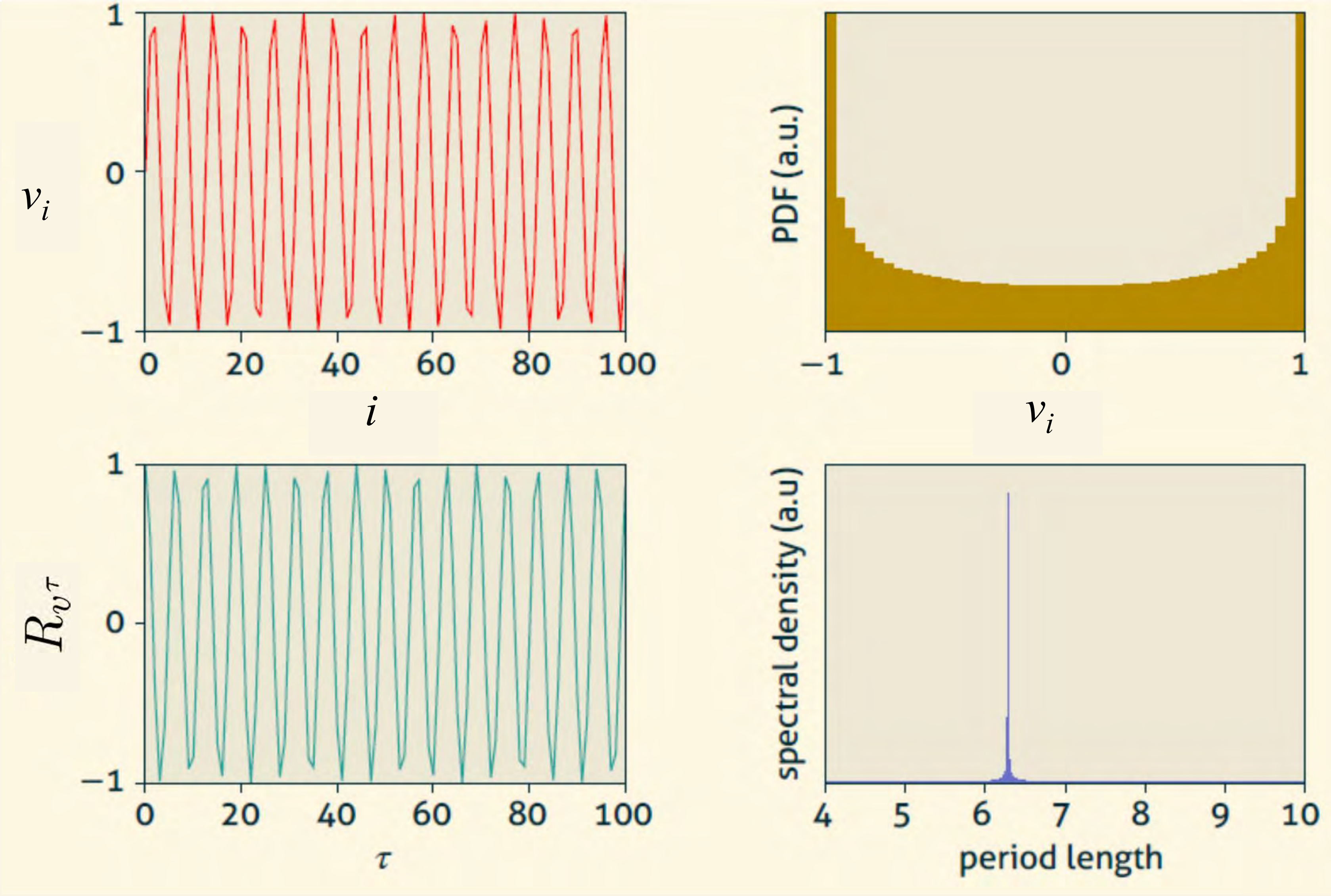

from: G. Ansmann

# Further Stochastic Processes

- continuous-time, e.g., stochastic differential equations

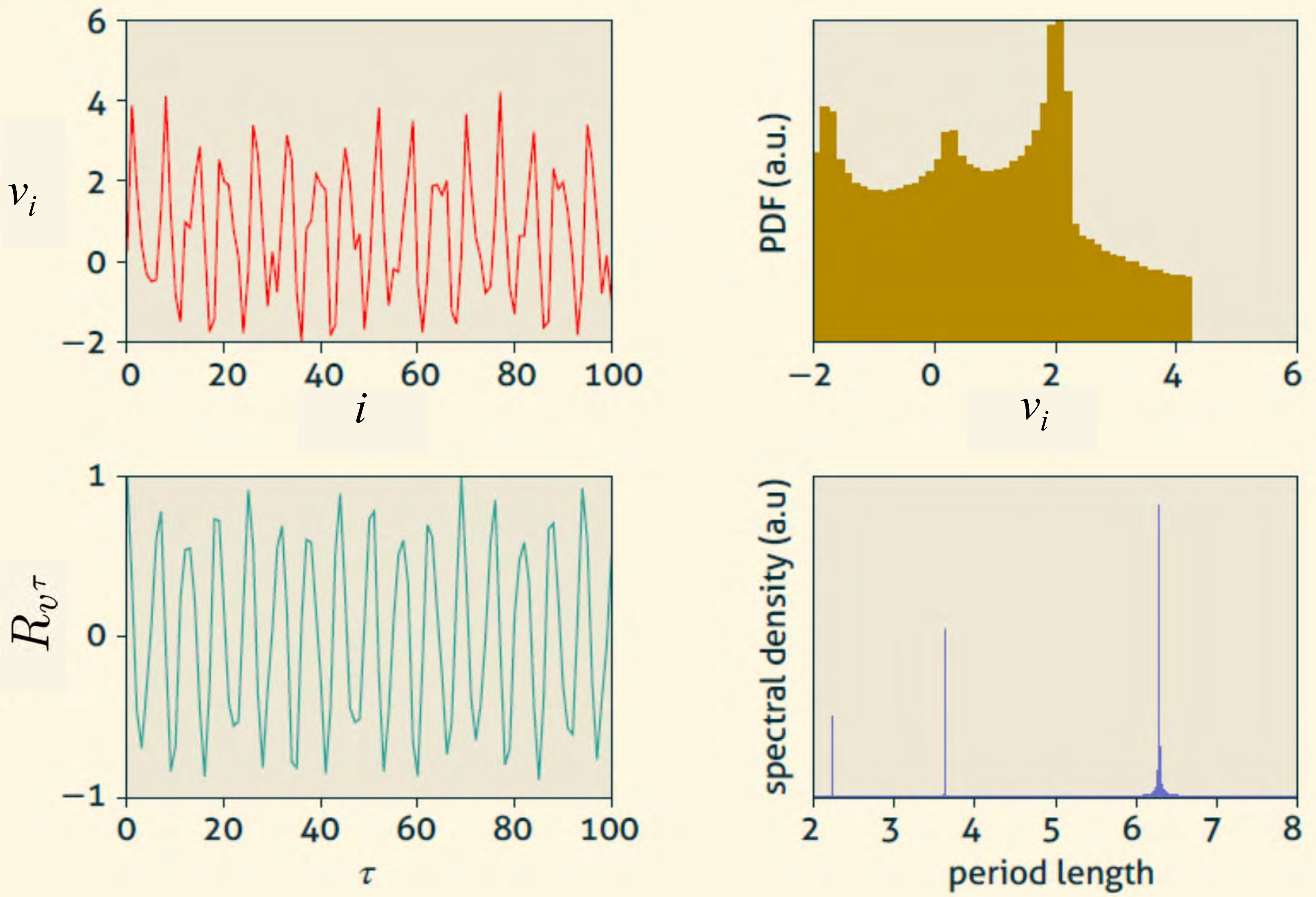
- nonlinear stochastic processes

# Applying linear methods — sine wave



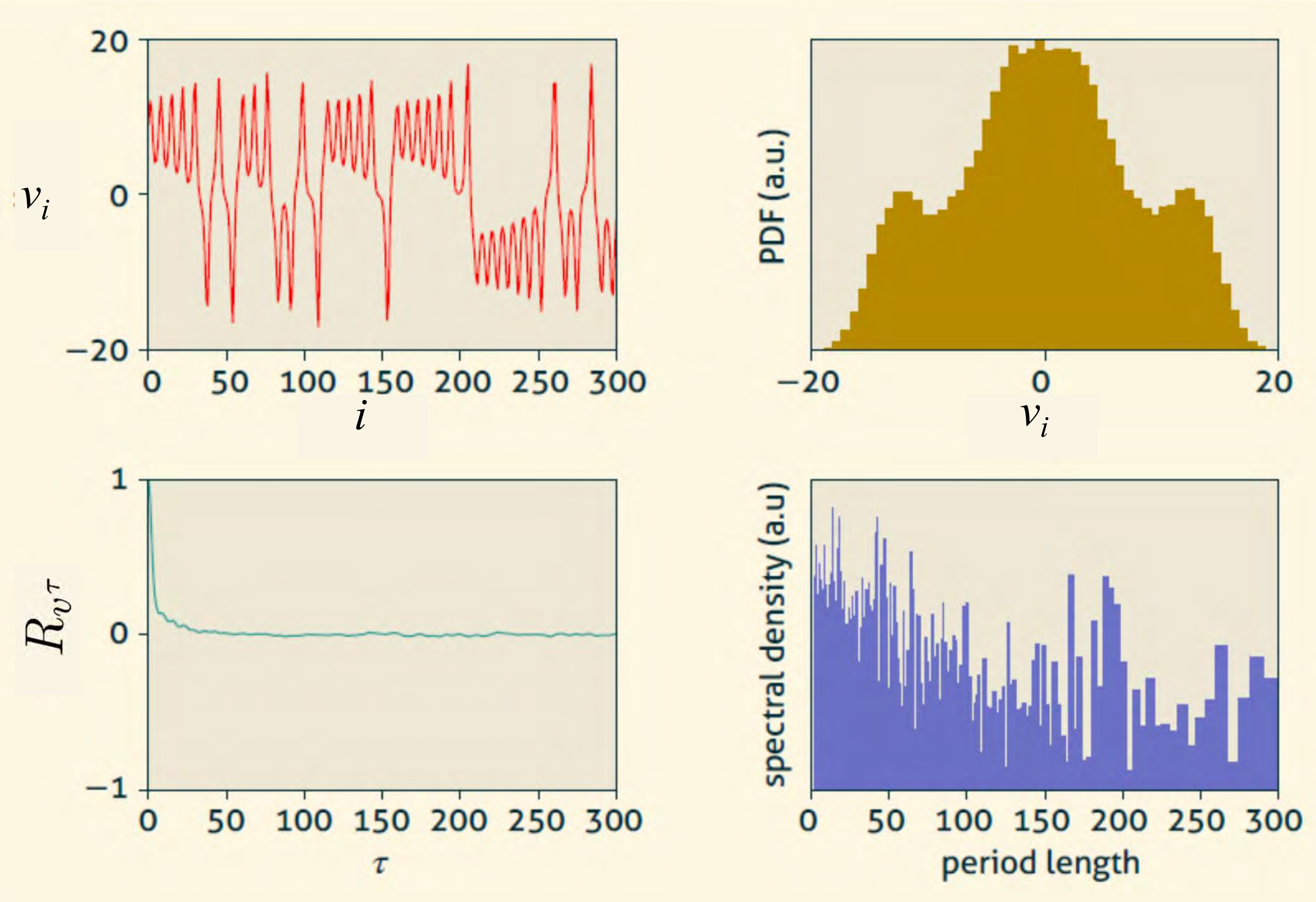

from: G. Ansmann

# Applying linear methods quasiperiodic



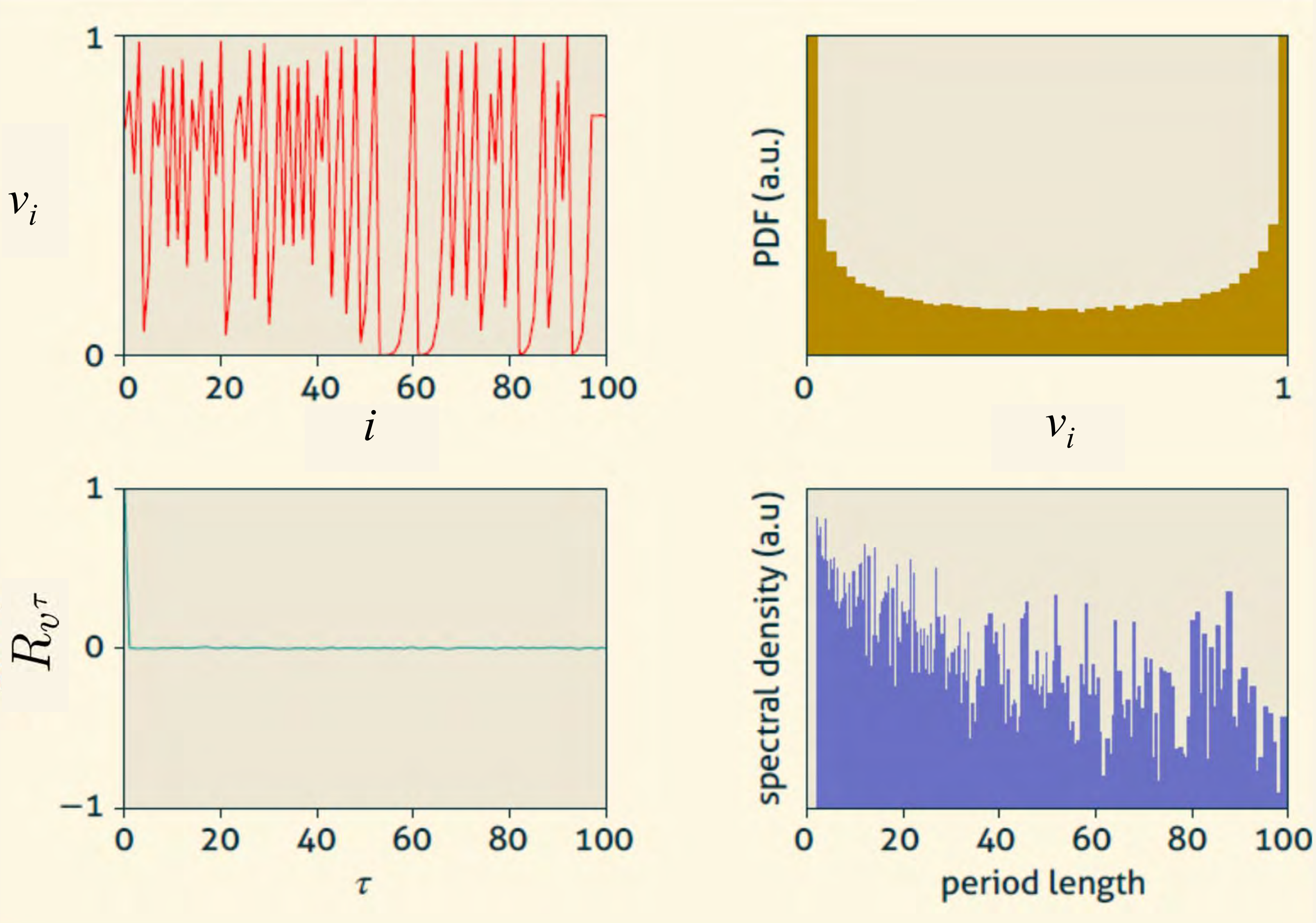

from: G. Ansmann

# Applying linear methods Lorenz oscillator



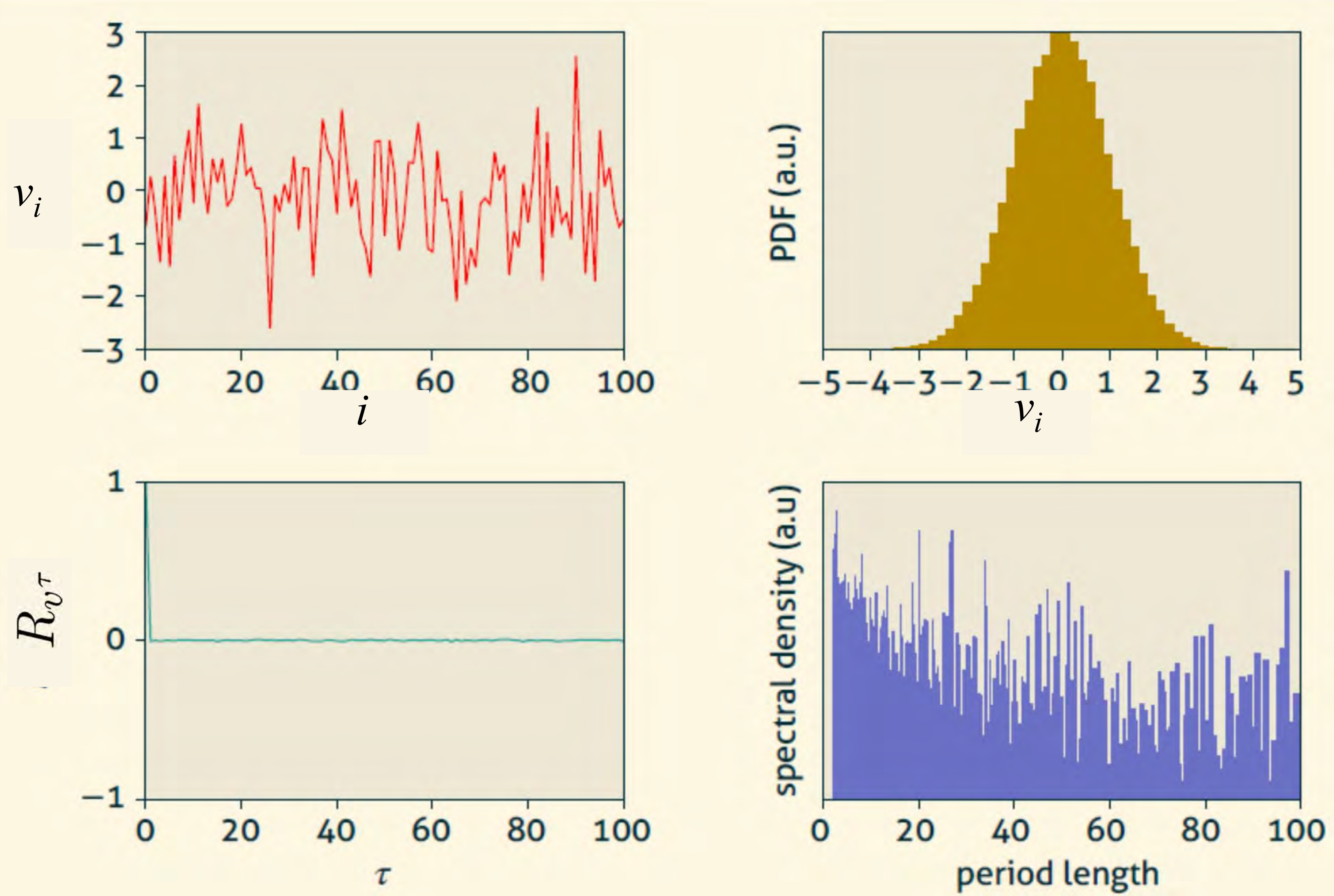

from: G. Ansmann

# Applying linear methods logistic map




from: G. Ansmann

## Applying linear methods Gaussian white noise




from: G. Ansmann

# Applying linear methods

***Capabilities:***

linear methods can:

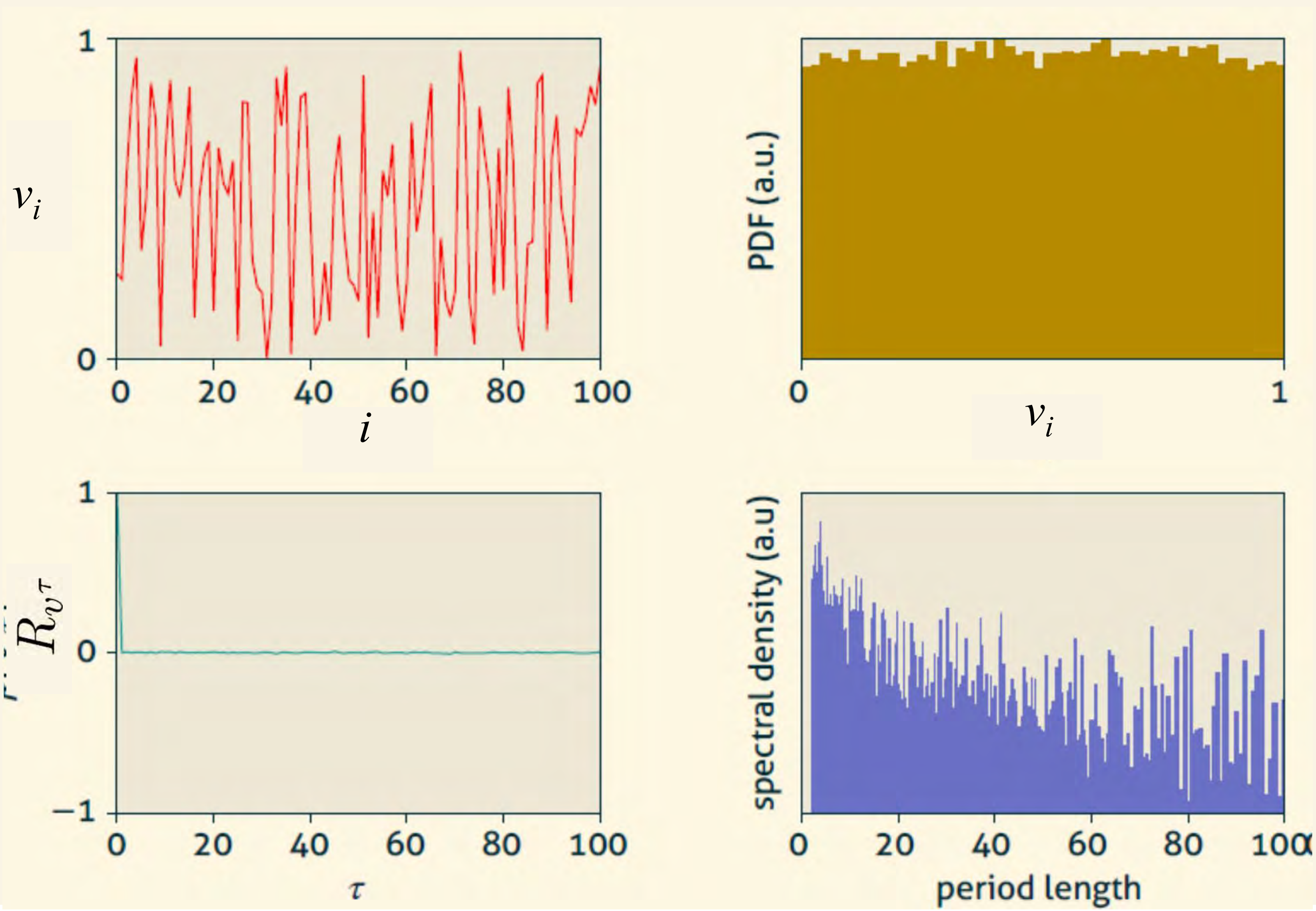
- detect periodic processes (non-decaying autocorrelation, discrete Fourier spectrum)
- hint at non-stochastic dynamics (not normally distributed)
- yield data-based, linear models that may not capture essential dynamical properties

***Restrictions:***

linear methods **cannot**:

- robustly distinguish noise from chaos
- yield nonlinear or chaotic models

# Applying linear methods Zaslavskii map




from: G. Ansmann

## Applying linear methods Zaslavskii map

a discrete-time dynamical system that maps a point ($x_n$, $y_n$) in the plane to a new point ($x_{n+1}$, $y_{n+1}$) :

$$
\begin{aligned}
x_{n+1} &= \Big(x_n + \nu\,(1 + \mu y_n) + \epsilon\nu\mu\cos\left(2\pi x_n\right)\Big) \bmod 1 \\
y_{n+1} &= \big(y_n + \epsilon\cos(2\pi x_n)\big)\exp(-\Gamma)
\end{aligned}
$$

where
$\Gamma = 3; \mu = \frac{1-\exp(-\Gamma)}{\Gamma}; \nu = \frac{400}{3}; \epsilon = 0.3$

*G.M. Zaslavskii (1978). "The simplest case of a strange attractor". Phys. Lett. A.* ***69*** *(3): 145–147*

## Applying linear methods Hénon map

a discrete-time dynamical system that maps a point $(x_n , y_n)$ in the plane to a new point $(x_{n+1} , y_{n+1})$ :

$$
\begin{aligned}
x_{n+1} &= 1 - a x_n^2 + y_n \\
y_{n+1} &= b x_n
\end{aligned}
$$

where
$a = 1.4; b = 0.3$

two-dimensional extension of logistic map

*M. Hénon (1976). "A two-dimensional mapping with a strange attractor". Comm. Math. Phys. 50 (1): 69–77.*

## Applying linear methods

# Stochasticity vs. Deterministic Chaos

- simple chaotic maps may be indistinguishable from stochastic processes with linear methods

- any pseudo-random-number generator is nothing but a very complex chaotic map

- but: nature may be more benign

*Any sufficiently complex determinism*
*is indistinguishable from stochasticity*