I. More on Omnigrams
II. ARMA Models

Jeffrey D. Scargle
NASA Ames Research Center (Retired)

AstroStat / CHASC January 28, 2026

* Review of Multiple Periodograms
* Autoregression
* Moving Averages
* The Wold Decomposition
* ARMA Models
* Generalized Deconvolution Example
* The Convolution Group
* The ARMA Convolution Group

Single Frequency Periodograms

$$y_n = a\ cos(\omega t_n - \theta) + b\ sin(\omega t_n - \theta)$$

$$\text{logL} = -\frac{1}{2}\sum_{n=1}^{N}\left(\frac{x_n - y_n}{\sigma_n}\right)^2 = -\frac{1}{2}\sum_{n=1}^{N} w_n\,(x_n - y_n)^2 \rightarrow -\frac{1}{2}\sum_{n=1}^{N}(x_n - y_n)^2$$

$$C = \sum_n w_n cos^2(\omega\ t_n - \theta)$$

$$S = \sum_n w_n sin^2(\omega\ t_n - \theta)$$

$$T = \sum_n w_n cos(\omega\ t_n - \theta) sin(\omega\ t_n - \theta$$

$$XC = \sum_n w_n x_n cos(\omega\ t_n - \theta)$$

$$XS = \sum_n w_n x_n sin(\omega\ t_n - \theta)\ .$$

$$P_{\text{Schuster}}(\omega) = |\sum x_n e^{-i\omega t_n}\ |^2$$
$$= (\sum x_n cos\ \omega t_n)^2 + (\sum x_n sin\ \omega t_n)^2$$

Table 14.1 *Periodograms*

Name	XC^2	$XCXS$	XS^2	a	b
S	1	0	1	-	-
SML	S/D	$-2T/D$	CC/D	(T XS-S XC)/D	(T XC-C XS)/D
LSP	$1/C$	0	1 /S	XC/C	XS/S
B	S/E	-2 T /E	C/E		
BLSP	$C^{-3/2}S^{-1/2}$	0	$C^{-3/2}S^{-3/2}$		

$$D = (T^2 - CS), E = D^{3/2}$$

In the periodogram names in this table S = Schuster, ML = max likelihood, B = Bayes/Bretthorst , LSP = Lomb-Scargle Periodogram.

Multiple Frequency Periodograms

$$\log L = -\frac{1}{2}\sum_{n=1}^{N}(\frac{x_n - y_n}{\sigma_n})^2 \ ,$$

$$y_n = a_1 \ cos(\omega_1 t_n - \theta) + b_1 \ sin(\omega_1 t_n - \theta)$$
$$+a_2 \ cos(\omega_2 t_n - \theta) + b_2 \ sin(\omega_2 t_n - \theta) \ ,$$

Least-Squares

$$R = a_1^2 \ CC11 + b_1^2 \ SS11 + a_2^2 \ CC22 + b_2^2 \ SS22$$
$$- 2(a_1 XC1 + b_1 XS1 + a_2 XC2 + b_2 XS2)$$
$$+ 2(a_1 b_1 CS11 + a_1 a_2 CC12 + a_1 b_2 CS12$$
$$+ b_1 a_2 SC12 + b_1 b_2 SS12 + a_2 b_2 CS22 \ )$$

Bayesian Marginalized Posterior

$$P(\omega_1, \omega_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} L(a_1, b_1, a_2, b_2) da_1 db_1 da_2 db_2 \ .$$

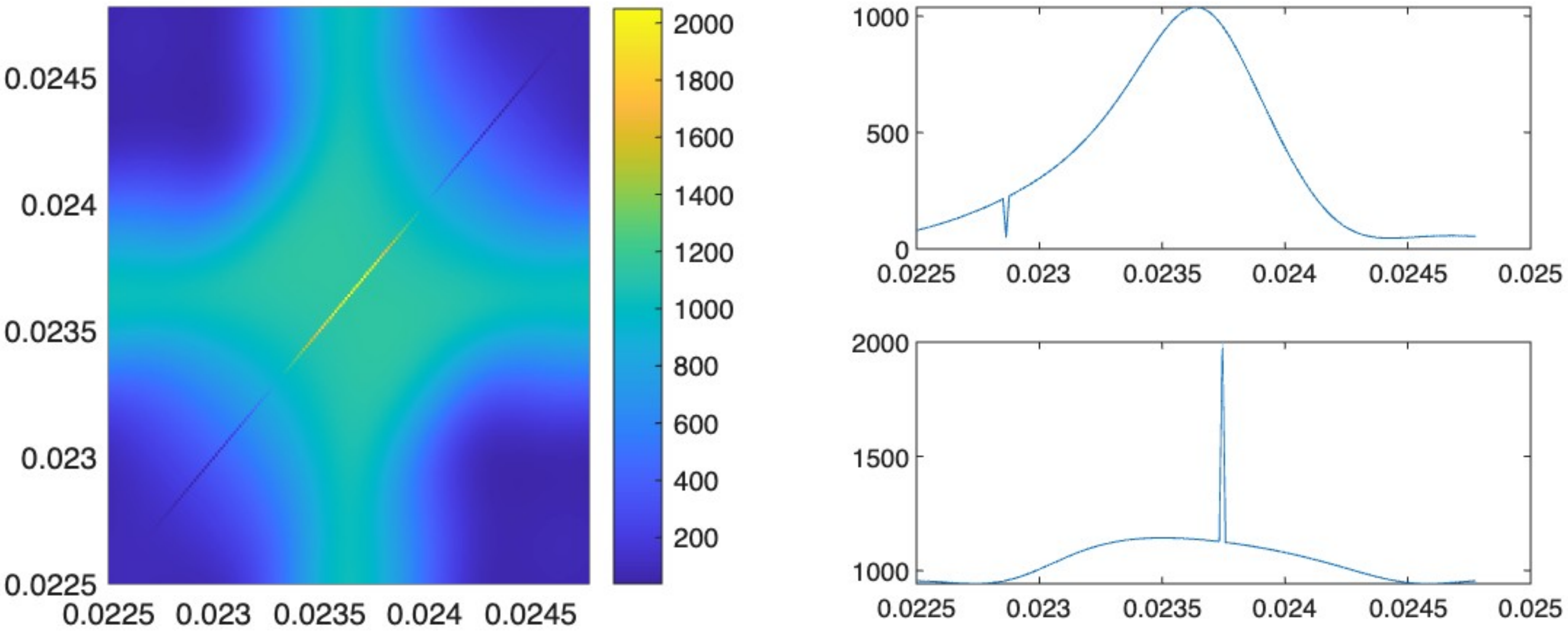
$$\int_{-\infty}^{\infty} e^{-(Ax^2+Bx+C)} \ dx = \sqrt{\frac{\pi}{A}} e^{(B^2/4A)-C} \ .$$

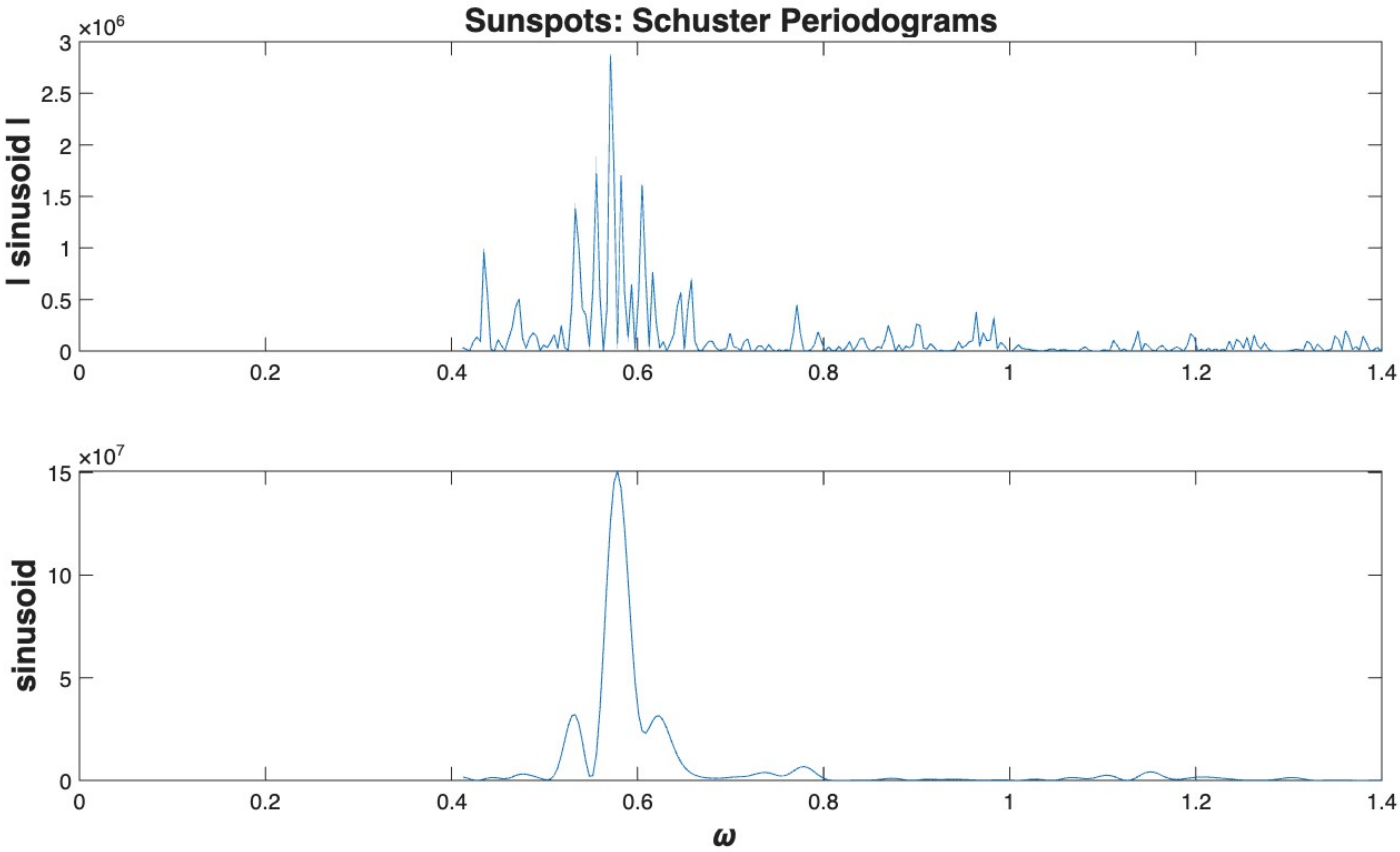
Other basis functions: just plug in!

$$y_n = a_1 \ cos((\omega_1 + \omega_2)t_n) \ cos((\omega_1 - \omega_2)t_n)$$
$$+b_1 \ sin((\omega_1 + \omega_2)t_n) \ cos((\omega_1 - \omega_2)t_n)$$
$$+a_2 \ cos((\omega_1 + \omega_2)t_n) \ sin((\omega_1 - \omega_2)t_n)/(\omega_2 - \omega_1)$$
$$+b_2 \ sin((\omega_1 + \omega_2)t_n) \ sin((\omega_1 - \omega_2)t_n)/(\omega_2 - \omega_1)$$

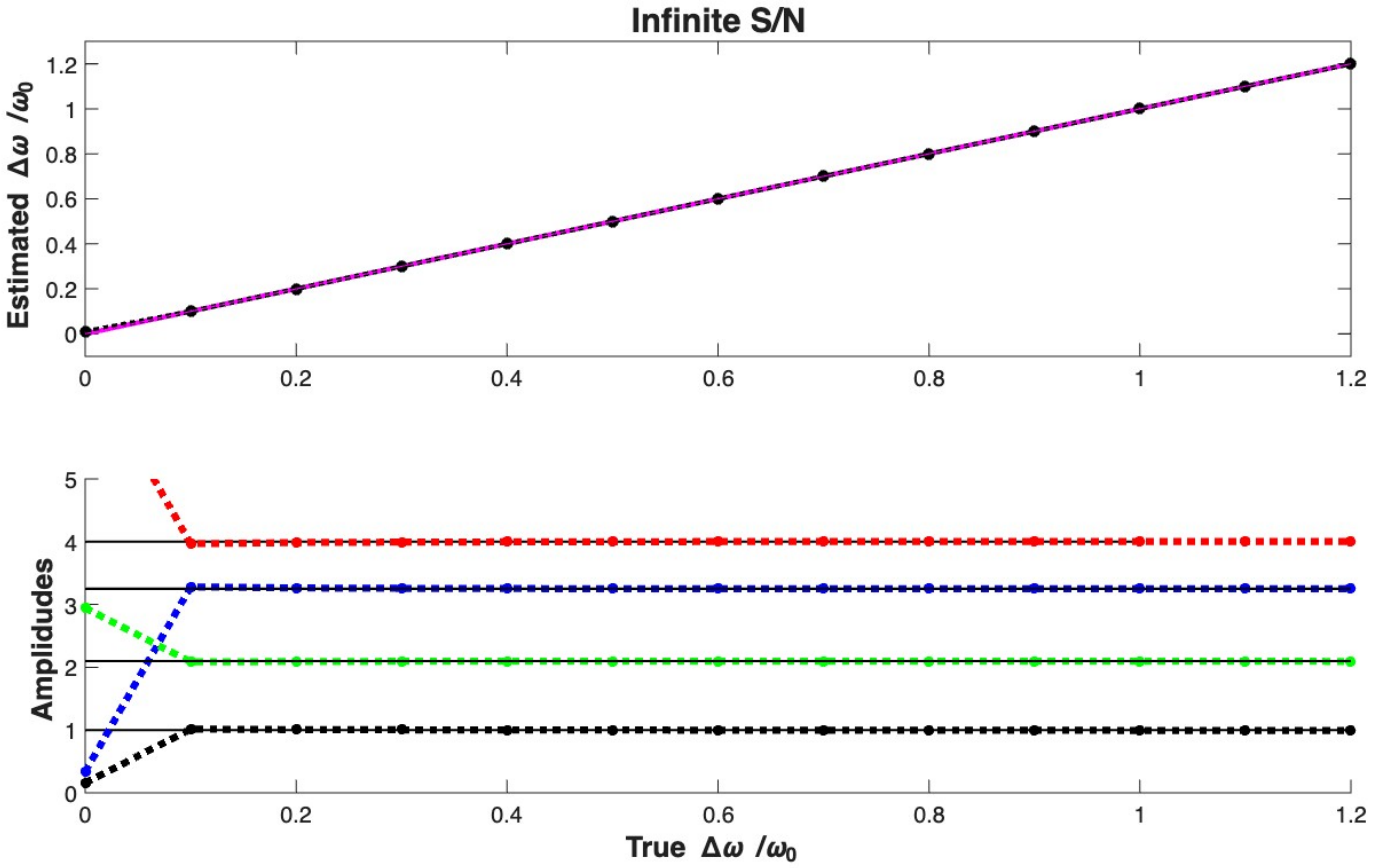
Baluev

The Diagonal Singulatity





Sunspots: Schuster Periodograms
×10^6
3
2.5
2
1.5
1
0.5
0
I sinusoid I
0
0.2
0.4
0.6
0.8
1
1.2
1.4
×10^7
15
10
5
0
sinusoid
ω



Infinite S/N
Estimated Δω /ω0
1.2
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
1.2
Amplitudes
5
4
3
2
1
0
True Δω /ω0

Omnigrams

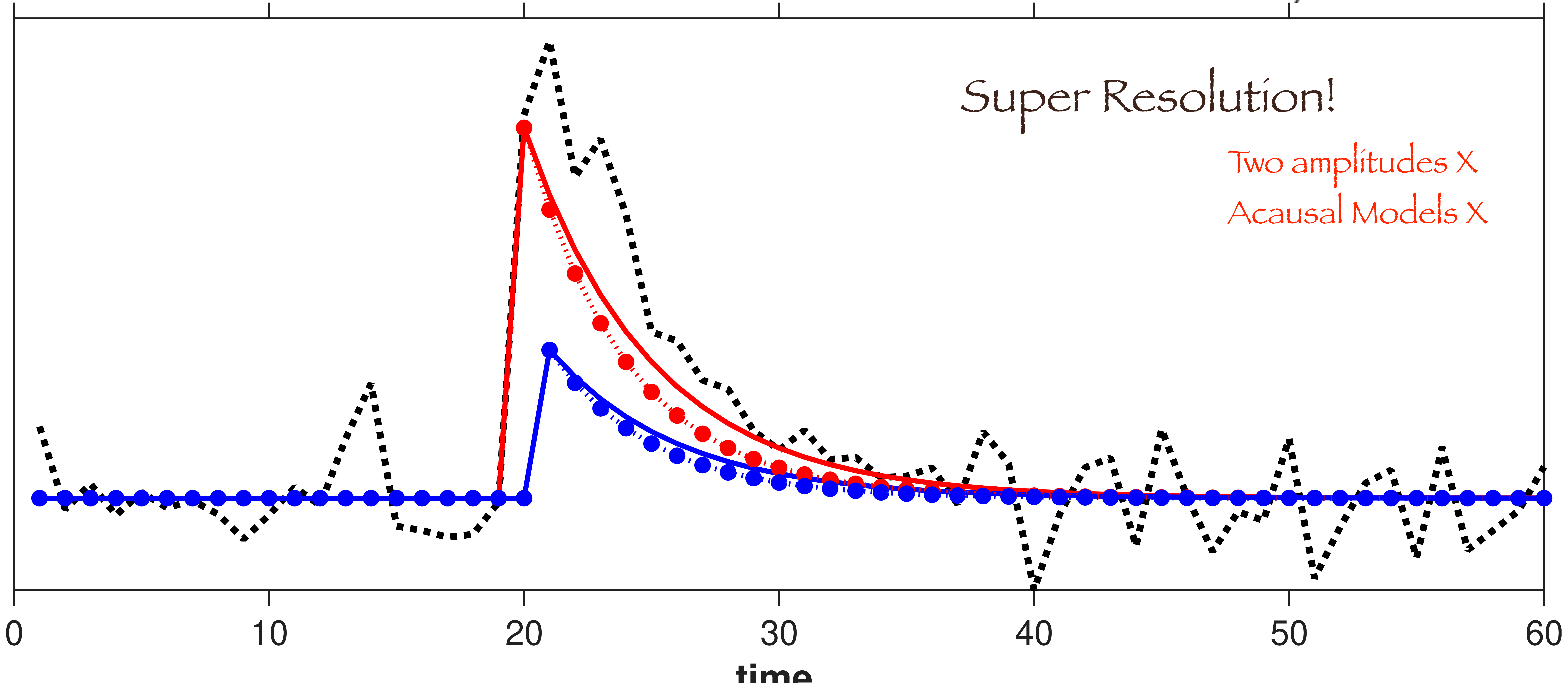
$$y_n = \sum_{j=1}^{m} B_j G_j(t_n, \hat{\omega}_j) + \epsilon_n \ .$$

$$-2\log L = \sum_{n=1}^{N} \left(\frac{x_n - \sum_{j=1}^{m} B_j G_j(t_n, \hat{\omega}_j)}{\sigma_n} \right)^2 \quad .$$

$$Q = \sum_{n=1}^{N} x_n^2 - 2 \sum_{j=1}^{m} B_j \sum_{n=1}^{N} x_n G_j(t_n, \hat{\omega}_j) \ + \ \sum_{n=1}^{N} [\ \sum_{j=1}^{m} B_j G_j(t_n, \hat{\omega}_j)\ ]^2 \ .$$

Three parameter time-domain Omnigram

$$F(t_n|\tau_{1,2}, a) \quad = e^{a(t_n - \tau_{1,2})} \qquad t_n \geq \tau_{1,2}$$
$$= 0 \qquad \text{otherwise,}$$



Autoregression: Remembrance of Things Past

$$x_n = a x_{n-1} + r_n$$

$$C = A^{-1} = \{1, a, a^2, a^3, \dots\}$$

Classically: |a| < 1 is necessary

A = (1, -a)

(Roots of characteristic equation outside unit circle in the complex plane: 1 - az = 0)

1 - az = 0 →

z = 1/a

$$AR(p): \qquad \sum_{k=0}^{p} A_k x_{n-k} = A * X = R$$

... consider a series of terms laid out in a horizontal line, such that each is derived from the preceding one according to a given law. Suppose this law takes the form of an equation of several consecutive terms and their indices indicating the position they occupy in the series. I call this the equation of finite first differences. The order or the degree of this equation is the difference between the indices of its first and last terms. In this way one can successively determine terms in the series, continuing indefinitely. But to do so it is necessary to know [initial] values in the series equal in number to to the degree of the equation. These values are the arbitrary constants giving the general term of the series, or of the integral of the difference equation.

$x_n = ax_{n-1} + ax_{n-2} \cdots$

AR(p)

... consider a series of terms laid out in a horizontal line, such that each is derived from the preceding one according to a given law. Suppose this law takes the form of an equation of several consecutive terms and their indices indicating the position they occupy in the series. I call this the equation of finite first differences. The order or the degree of this equation is the difference between the indices of its first and last terms. In this way one can successively determine terms in the series, continuing indefinitely. But to do so it is necessary to know [initial] values in the series equal in number to to the degree of the equation. These values are the arbitrary constants giving the general term of the series, or of the integral of the difference equation.

$x_n = ax_{n-1} + ax_{n-2} \cdots$

AR(p)

Pierre Simon de LaPlace (1749-1827),
in Essai philosophique sur les probabiliti´es

Autoregression is not Markov!

AR(1)'s memory one time-step into the past is similar to the defining property of a Markov process (Section 21.4) as one whose

probability distribution depends on previous states $x_{n-k}, k > 0$

In contrast to such probabilistic memory, the dependence on the past in Equation (19.14) is deterministic. The innovation, the only probabilistic element, has no memory of the past.

- Moving Average Models

All models are wrong; some are useful.

(Origin unknown, but sometimes attributed to George E. P. Box.)

$$X = C * R, \quad \text{i.e.} \quad x_n = \sum_{k=0}^{p} C_k r_{n-k}$$

n = 0 is the "origin of time"

Causality means $C_k = 0$ for $k < 0$

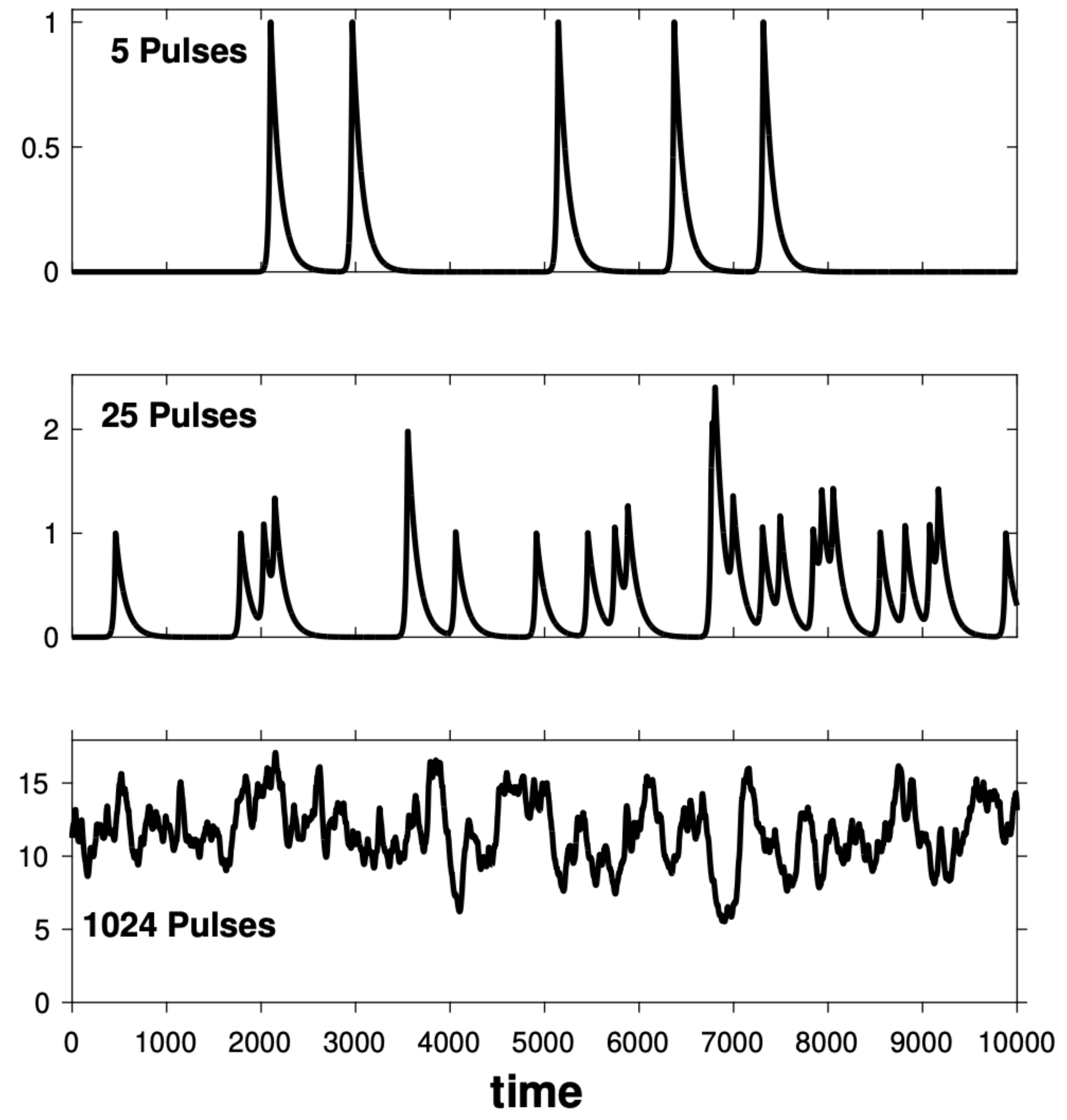


Figure 19.1 **The March From Discrete Events to Gaussianity.** These time

In the limit of a large degree of pulse overlap, the process becomes Gaussian and is fully described by its second order statistics. Much information about pulse shapes is thereby irretrievable lost.

One example: *fast-rise/exponential-decay pulses* (FREDs) cannot be distinguished from their time reverse. Due to the loss of phase information in this limit, the true pulse shape cannot be recovered by any method – including those introduced in Section **??**.

The Wold Decomposition

The Wold Decomposition Theorem: Any stationary process X can be decomposed into the sum of two processes: one purely deterministic, and the other a moving average consisting of the convolution of a constant function (filter C) with an uncorrelated Gaussian process (innovation R):

$$X = C * R + D \qquad x_n = \{\sum_{k \geq 0} C_k R_{n-k} + D_n\} \tag{19.84}$$

Seriously equal

Constant

Random

Deterministic

Chatfield (2004) considers it mainly of theoretical interest, and has rarely found it of much assistance.

On the other hand, in his review of the book Wold (1938), Neyman (1939) said the Wold Theorem is of considerable interest in describing the structure of the most general discrete stationary, and recommended the book to time series analysts in economics and statistics, for a practically useful description of the theory.

The Wold Decomposition Theorem: Any stationary process X can be decomposed into the sum of two processes: one purely deterministic, and the other a moving average consisting of the convolution of a constant function (filter C) with an uncorrelated Gaussian process (innovation R):

$$X = C * R + D \qquad x_n = \{\sum_{k \geq 0} C_k R_{n-k} + D_n\} \tag{19.84}$$

Explicit Linearity The Wold Decomposition is linear in several ways:

- Process $\mathbf{D}$ is linearly deterministic (but not necessarily linear itself).
- The moving average $\mathbf{C} * \mathbf{R}$ and trend $\mathbf{D}$ are simply added together.
- The MA expresses linear response to the innovation.

Stationarity is a very powerful condition!

The Wold Decomposition Theorem: Any stationary process X can be decomposed into the sum of two processes: one purely deterministic, and the other a moving average consisting of the convolution of a constant function (filter C) with an uncorrelated Gaussian process (innovation R):

$$X = C * R + D \qquad x_n = \{\sum_{k\geq 0} C_k R_{n-k} + D_n\} \tag{19.84}$$

Filter Specificity The linear form of the decomposition is rather special itself, but there is more to come! The MA filter has these specific properties – consequences of the theorem following from the single assumption of stationarity, and not imposed constraints:

(a) C is *causal.*
(b) C is *constant.*
(c) C is *minimum delay*

Problems with Application to Real Data

Note: The distribution $R \sim N(0, \sigma)$ of the Wold innovation may be astrophysically inappropriate in at least two ways. Its zero mean is not consistent with non-negativity of outbursts of radiation. The shape of the normal distribution, with its finite probability of unbounded negative values, may differ from the actual distribution.

Causality Need Not Apply

- For sequential data in domains other than time (spatial, wavelength, etc.) the concepts replacing *past, present* and *future* do not have the same cause-and-effect type significance.
- We will see in Section 19.2 that for some filters the time at which input begins is ambiguous. It is then natural to model response both before and after the fiducial input moment (Chapter 20).
- In retrospective analysis, measurements over the full observation interval are available to the analyst. Hence there is no reason why the modeling the value at a given point of time can't usefully incorporate values in that point's future.

In at least these three major contexts, the conventional restriction to causal models can result in underutilization of information in the data.

$$
\begin{aligned}
X(t) &= x_0 \, e^{-c(growth) \, |t-t_0|} && t \leq t_0 \\
&= x_0 \, e^{-c(\;\;decay) \, |t-t_0|} && t \geq t_0
\end{aligned}
$$

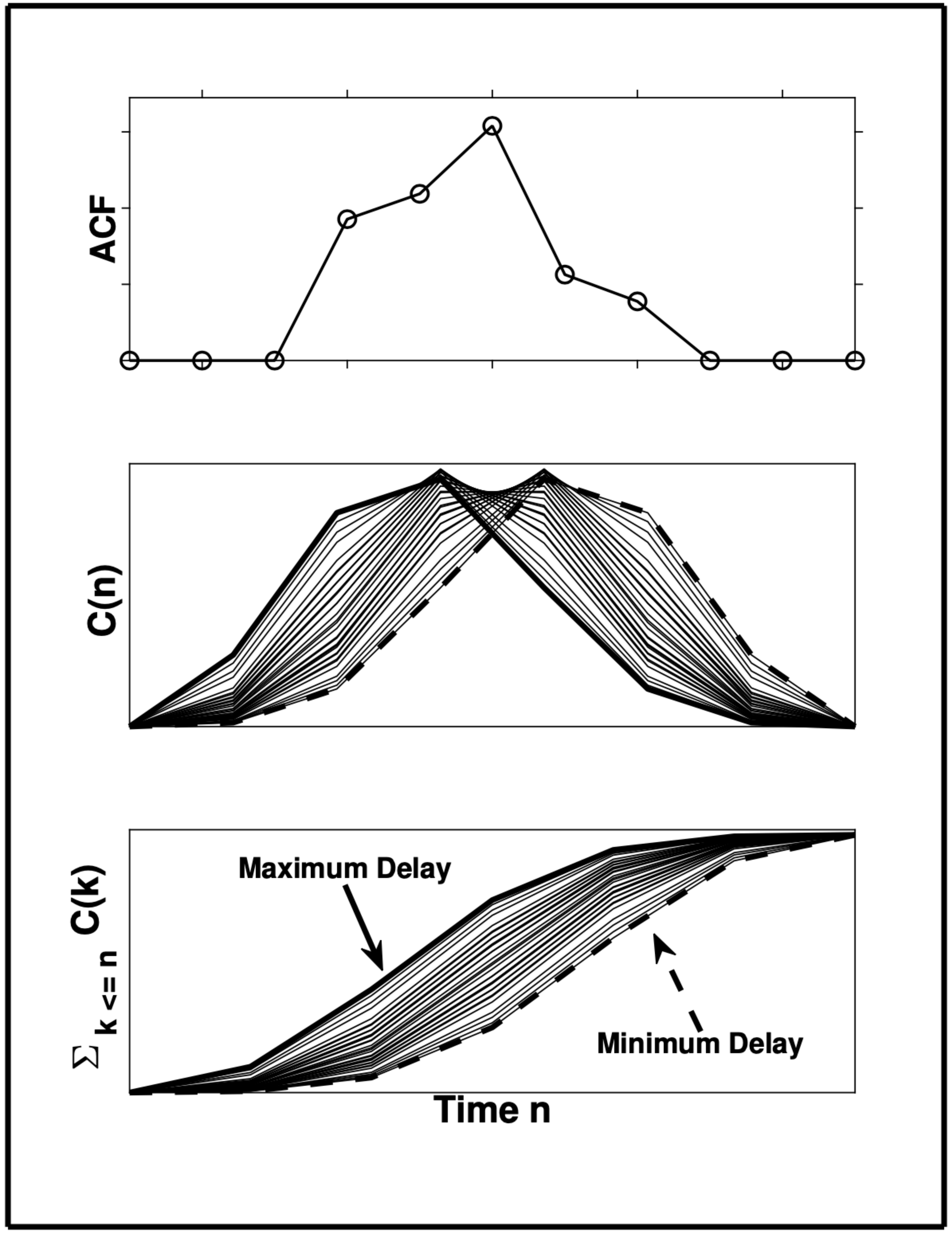
The Arrow of Time

Analysis methods limited to second-order statistics do not access information needed to distinguish between minimum and maximum delay filter shapes. Therefore they cannot capture the *arrow of time* – to tell whether the "movie" is running in reverse or not.

In failing to distinguish any of the filters of intermediate delay character, these methods miss more than just the arrow of time.

And what is this Minimum Delay Thing?

Delay Properties of Moving Average Filters



Extended Wold Decomposition:

Given a stationary process X, there exist:

1 A linearly deterministic process D
2 A family of uncorrelated, zero-mean random processes $\{R^k\}$
3 A family of (two-sided) moving average filters $\{C^k\}$

such that

$$X = C^k * R^k + D \ , \tag{20.1}$$

for all k. The filter family is the set of all those that have the same autocorrelation function as X: one minimum delay, one maximum delay, and the rest mixed delay.

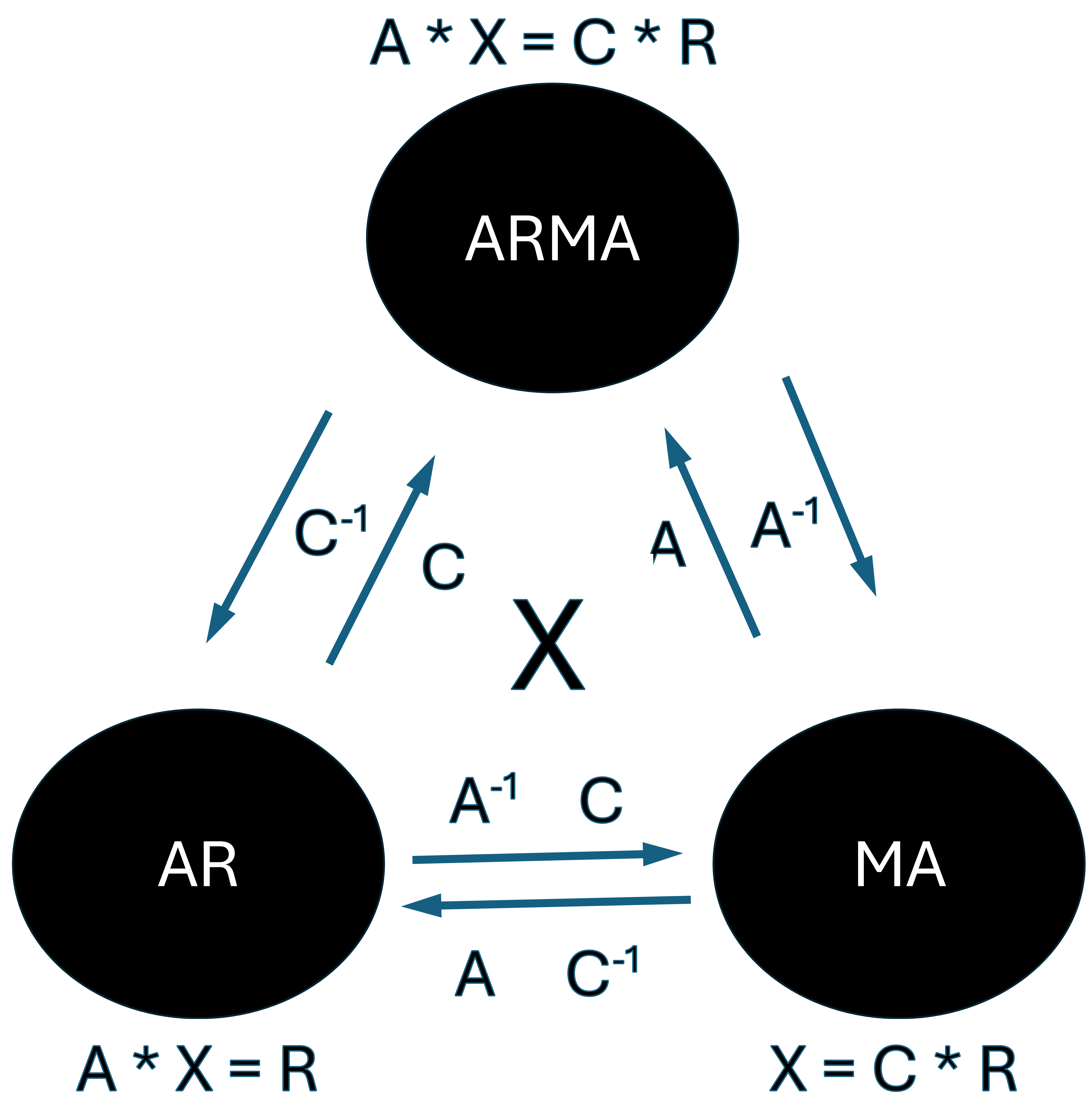
- ARMA Models

$$x_n = \sum_{k=0}^{p} c_k r_{n-k} + \sum_{k=1}^{q} a_k x_{n-k}$$

Explicit Definition	$\sum_{k=0}^{q} a_k x_{n-k} = \sum_{k=0}^{p} c_k r_{n-k}$
Convolutional	$A * X = C * R$
Fourier Transform	$FT_A(k)\ FT_X(k) = FT_C(k)\ FT_R(k)$
Z-Transform	$Z_A(z)\ Z_X(z) = Z_C(z)\ Z_R(z)$
Laplace Transform	$(\sum_k a_k e^{-ks})\ \mathcal{L}_X(s) = (\sum_k c_k e^{-ks})\ \mathcal{L}_R(z)$
Shift Operator	$\phi(\mathbf{B})x_n = \theta(\mathbf{B})r_n$
Characteristic Polynomial	$\phi(z)x_n = \theta(z)r_n$
Factored Characteristic Polynomial	$\prod_k(1 - z\alpha_k)x_n = \prod_k(1 - z\gamma_k)r_n$

$$z \leftrightarrow e^{-i\omega} \leftrightarrow B \leftrightarrow e^{-s}$$

The ARMA Convolution Group



$$X * Y = \mathcal{F}^{-1}[\,\mathcal{F}(X)\mathcal{F}(Y)\,]\,,$$

$$A^{-1} = \mathcal{F}^{-1}[\,1/\mathcal{F}(A)\,]$$

- Deconvolution Example

Conjecture:

Given: Stationary process $X = \{x_n\}$.

Define:

- Ω, an index set (positive or negative, non-zero integers)
- A linear regression form:

$$\hat{x}_n = \sum_{k \in \Omega} B_k x_{n-k} \quad , \tag{20.3}$$

  where $\{B_k, k \in \Omega\}$ are unknown real coefficients.
- A measure of the distance between this model and X:

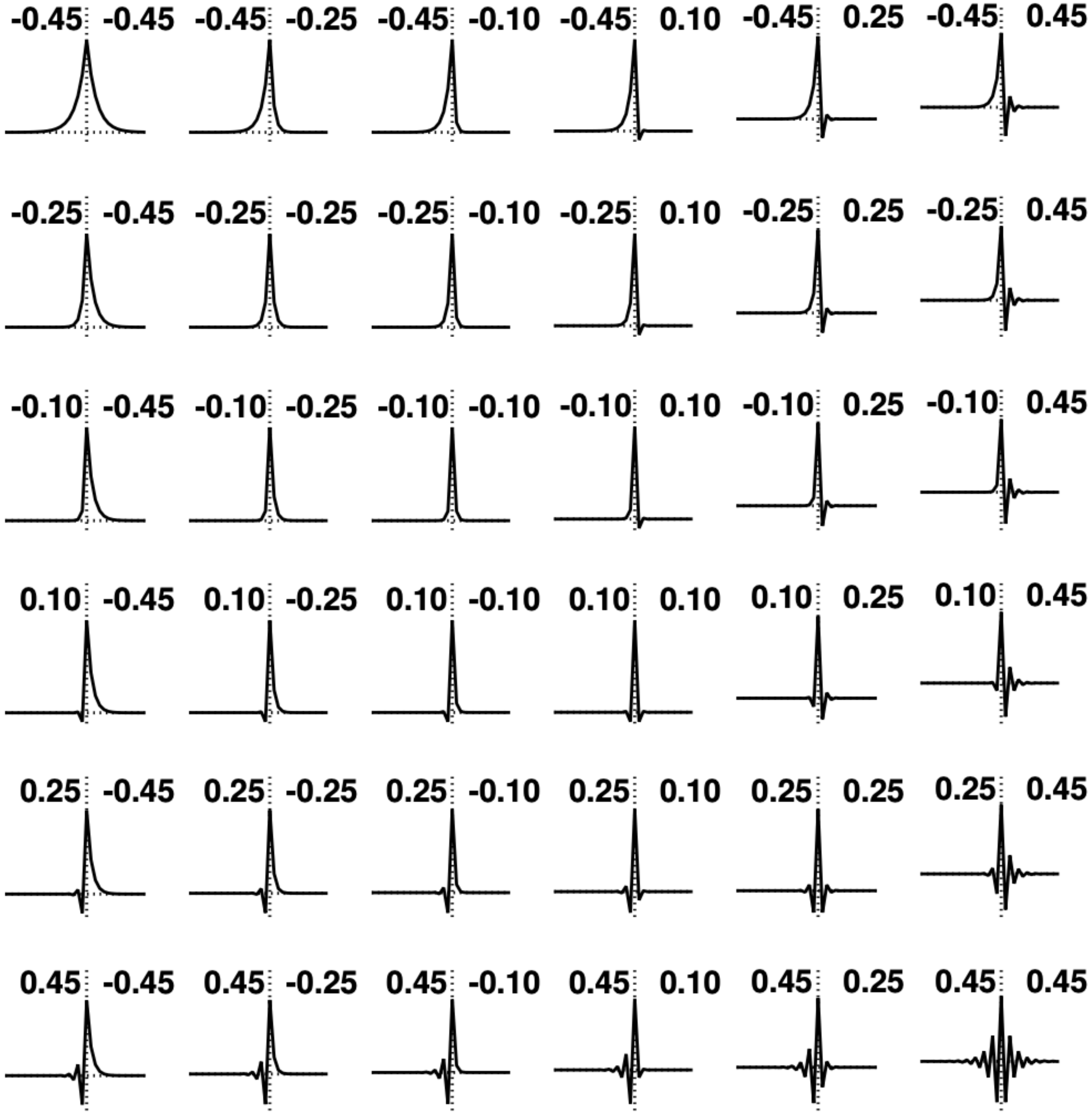
$$D(X|\hat{X}) = E_n[F(X_n, \hat{x}_n)] \tag{20.4}$$

  for some function F and estimator E_n (e.g. some kind of average).

If a minimum of $D(X|\hat{X})$ with respect to the parameters B_k exists, Equation (20.3) evaluated with the minimizing B_k values is an exact representation of process X.

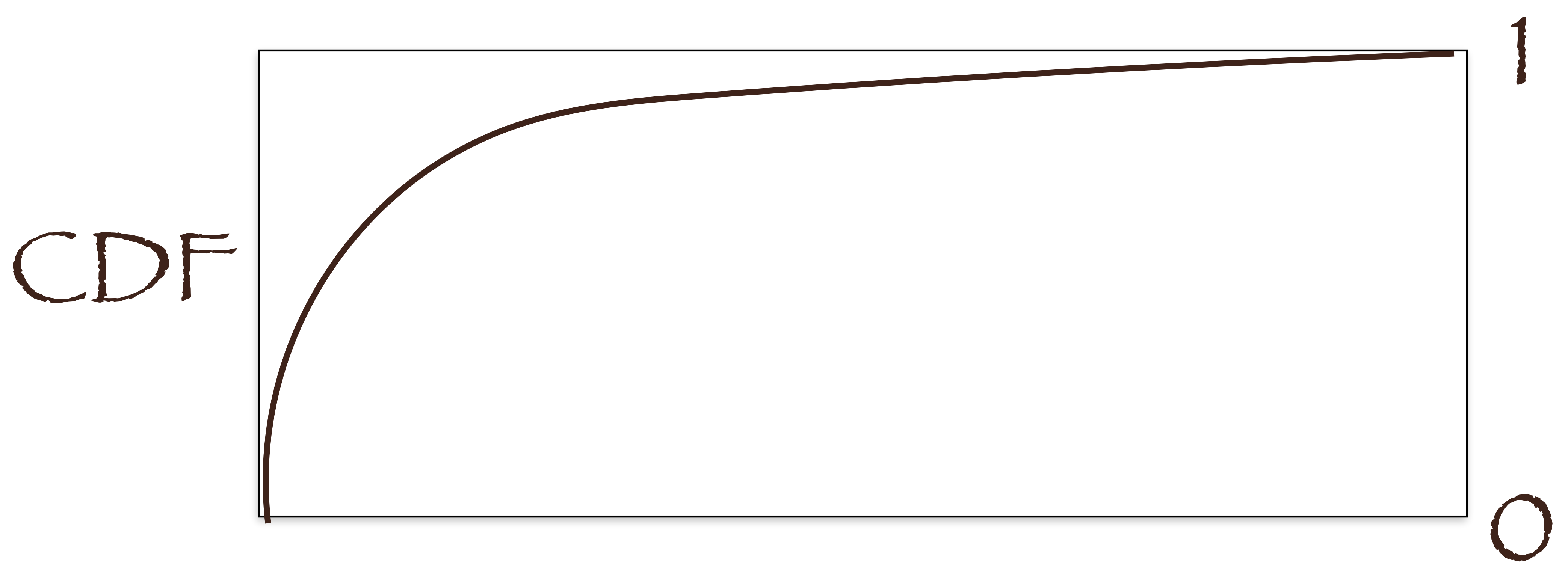
$$\sum_{k=-q_{\text{future}}}^{q_{\text{past}}} a_k x_{n-k} = \sum_{k=-p_{\text{future}}}^{p_{\text{past}}} c_k r_{n-k}$$

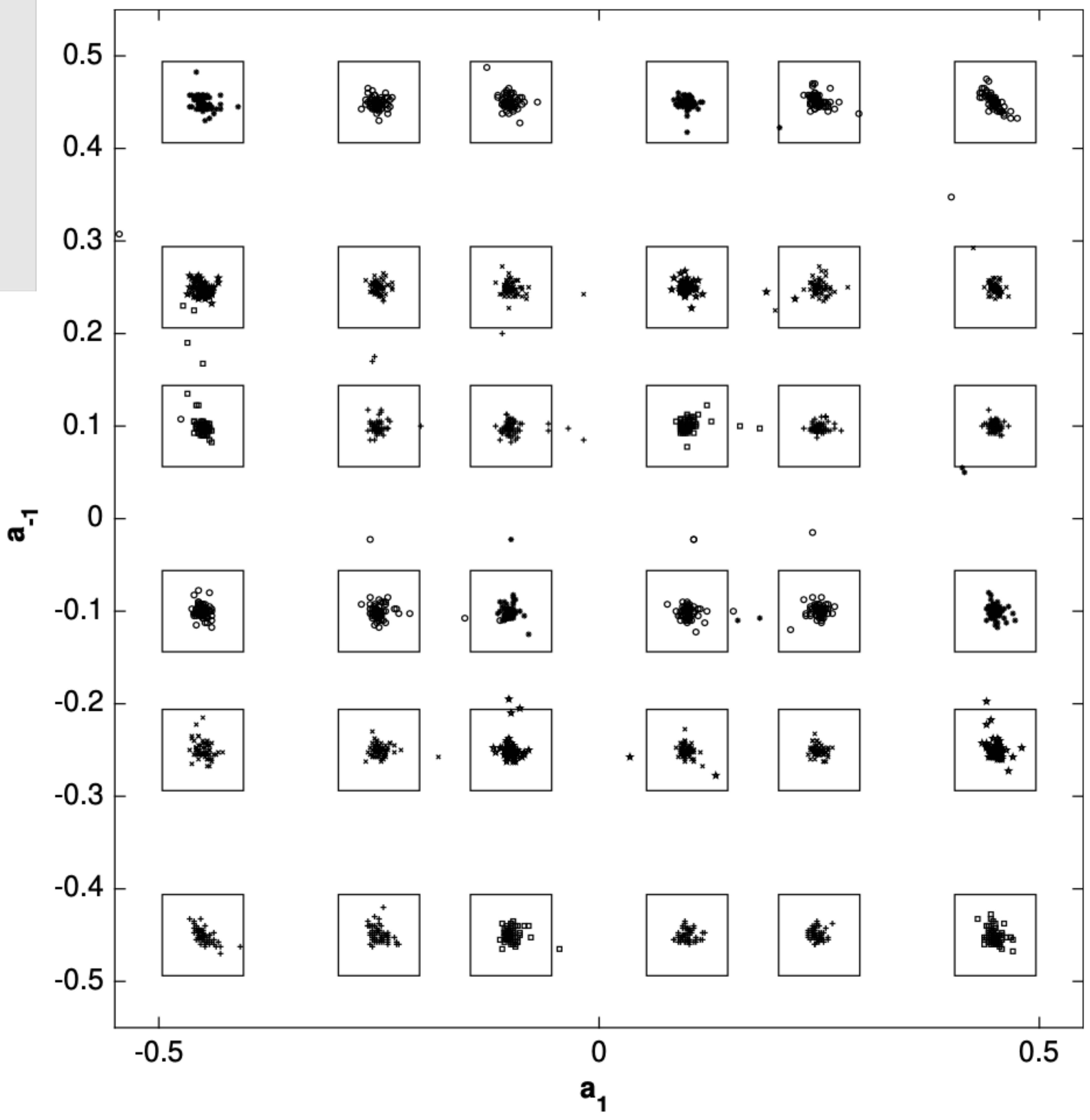
$$x_n = a_{-1}x_{n-1} + a_1 x_{n+1} + R_n$$



Pseudocode: **A Distributional Sparseness Measure**

```
Input: array R of length nn
CDF = (1: nn-1)/(nn - 1)
R = R - min(R);R = R / max(R)    % Normalized interval
R = sort( R )
sparse_metric = sum( diff( R ) * CDF )
```





The Convolution Group

The ARMA Convolution Group

Properties of Convolution

1. **Closure:** The convolution of two filters is a filter.
2. **Commutativity:** $A * B = B * A$
3. **Associativity:** $A * (B * C) = (A * B) * C$
4. **Index swap:** $\sum_k A_k B_{n-k} = \sum_{k'} A_{n-k'} B_{k'}$
5. **Length:** Length(C) = Length(A) + Length(B) - 1
8. **Polynomials:** Polynomial Multiplication = Filter Convolution
9. **Identity:** There is a filter I such that $I * C = C$ (δ-function).

$$X * Y = \mathcal{F}^{-1}[\,\mathcal{F}(X)\mathcal{F}(Y)\,]\,,$$

$$A^{-1} = \mathcal{F}^{-1}[\,1/\mathcal{F}(A)\,]$$

$$S = \{\mathbf{A_0}, A_1, A_2, \ldots A_{M-1} \mid A_0 = 1\,,\;\; F(A)\big|_{k=0}^{M-1} \neq 0\}$$

Existence of convergent Inverses is the only nontrivial group property

A commutative (aka Abelian) group is a set $\mathcal{S} = \{C_n\}$ of things C_n (elements) and an operator $\oslash$ applied to pairs of things, that satisfy the following for all such elements:

Closure: $C_1 \oslash C_2$ is in S
Commutativity: $C_1 \oslash C_2 = C_2 \oslash C_1$
Associativity: $C_1 \oslash (C_2 \oslash C_3) = (C_1 \oslash C_2) \oslash C_3$
Identity: There is an element I in S such that $I \oslash C = C$
Inverse: There is an element C^{-1} such that $C^{-1} \oslash C = I$

Fix length of filters + Fourier wraparound

The ARMA Convolution Group

$$\mathcal{M} = (C, A)$$

To make a group:

(a) the *group operator* $\oslash$, by specifying $\mathcal{M}_1 \oslash \mathcal{M}_2$
(b) the *identity element* $\mathcal{I}$, satisfying $\mathcal{I} \oslash \mathcal{M} = \mathcal{M}$
(c) the *inverse* $\mathcal{M}^{-1} = \oslash^{-1} \mathcal{M}$ satisfying $\mathcal{M}^{-1} \oslash \mathcal{M} = \mathcal{I}$

$$\mathcal{M}_1 \oslash \mathcal{M}_2 = (C_1 * C_2, A_1 * A_2)$$

$$\mathcal{M}^{-1}(C, A) = (A^{-1}, C^{-1})$$

$$\mathcal{M}^{-1}(\delta, A) = (A^{-1}, \delta)$$

$$\mathcal{M}^{-1}(C, \delta) = (\delta, C^{-1})$$

The Fundamental Theorem of Convolution

Any filter A of length M is the convolution of M elementary dipole filters C_k, each of length 2:

$$A = A_1 * A_2 * A_3 * \cdots * A_M \qquad (19.49)$$

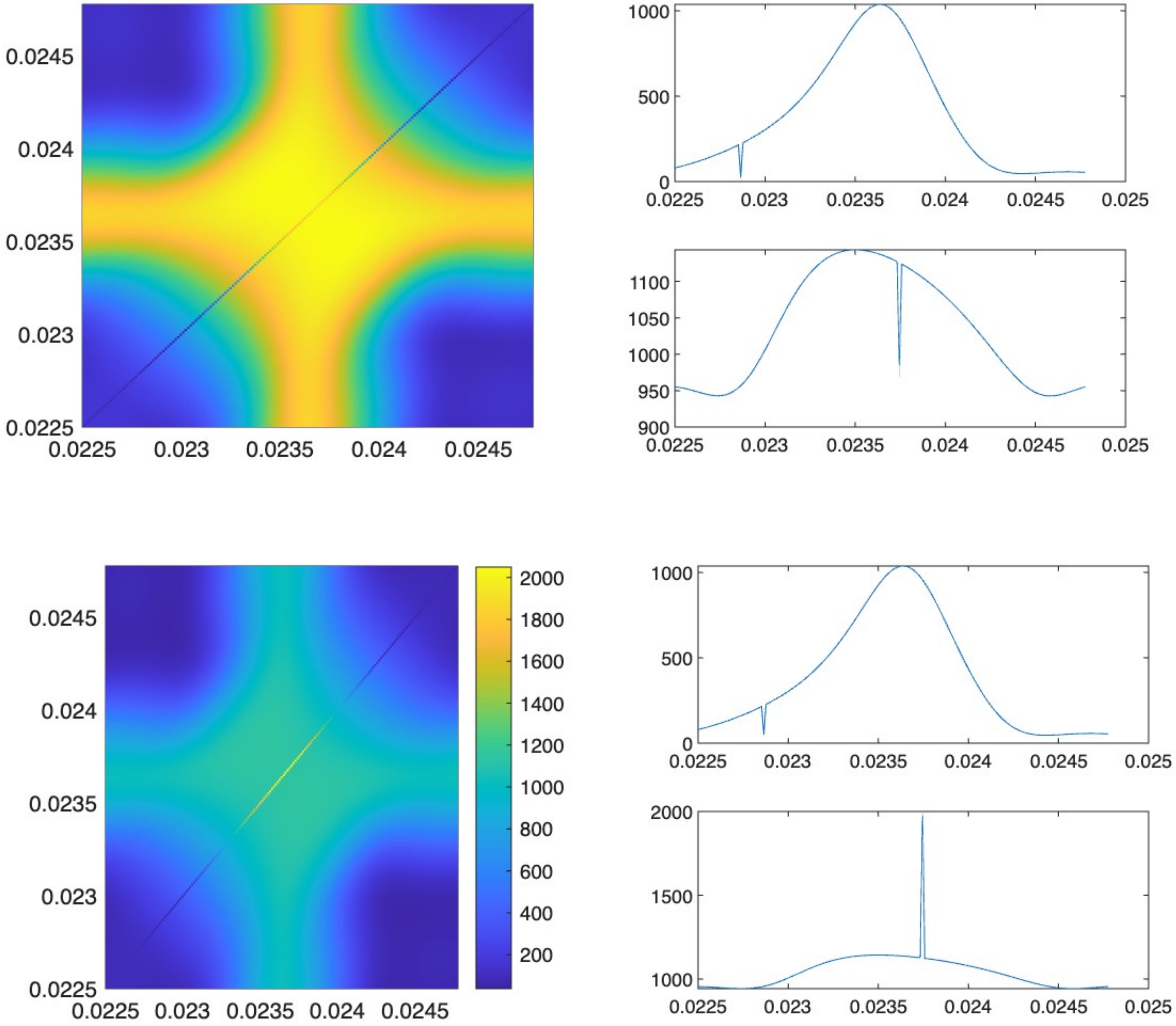
where $A_k = \{\mathbf{1}, a_k\}$.
By the same token, an M-th order polynomial can be factored into binomials:

$$\phi(z) = c_0 \prod_{k=1}^{M} (1 - c_k) \qquad (19.50)$$

Group generators: dipoles {1,-a }

$$\prod_k (1 - z\alpha_k) x_n = \prod_k (1 - z\gamma_k) r_n$$

- Backup


0.0245
0.024
0.0235
0.023
0.0225
0.0225 0.023 0.0235 0.024 0.0245
1000
500
0
0.0225 0.023 0.0235 0.024 0.0245 0.025
1100
1050
1000
950
900
0.0225 0.023 0.0235 0.024 0.0245 0.025
2000
1800
1600
1400
1200
1000
800
600
400
200
2000
1500
1000