Discrete Chi-Square Method can model and forecast complex time series

Lauri Jetsu
lauri.jetsu@helsinki.fi
Department of Physics, P.O. Box 64,
FI-00014 University of Helsinki, Finland

Abstract

We show how intensive, large and accurate time series can allow us to see through time. Many phenomena have aperiodic and periodic components. An ideal time series analysis method would detect such trend and signal(-s) combinations. The widely-used Discrete Fourier Transform (DFT) and other frequency-domain parametric time series analysis methods have many application limitations constraining the trend and signal(-s) detection. We show that none of those limitations constrains our Discrete Chi-square Method (DCM) which can detect signal(-s) superimposed on an unknown trend. Our simulated time series analyses ascertain the revolutionary Window Dimension Effect (WD-effect): *"For any sample window* $\boldsymbol{\Delta T}$*, DCM inevitably detects the correct* $\boldsymbol{p(t)}$ *trend and* $\boldsymbol{h(t)}$ *signal(-s) when the sample size* $\boldsymbol{n}$ *and/or data accuracy* $\boldsymbol{\sigma}$ *increase."* The simulations also expose the DFT's weaknesses and the DCM's efficiency. The DCM's backbone is the Gauß-Markov theorem that the Least Squares (LS) is the best unbiased estimator for linear regression models. DCM can not fail because this simple method is based on the computation of a massive number of linear model LS fits. The Fisher-test gives the signal significance estimates and identifies the best DCM model from all alternative tested DCM models. The analytical solution for the non-linear DCM model is an ill-posed problem. We present a computational well-posed solution. The DCM can forecast complex time series. The best DCM model must be correct if it passes our Forecast-test. Our DCM is ideal for forecasting because its WD-effect spearhead is robust against short sample windows and complex time series. In our appendix, we show that DCM can model and forecast El Niño.

Keywords: Time series analysis, Non-linear and non-stationary processes, Ill-posed problems, Computational statistics, Forecasting, Discrete Chi-square Method, Discrete Fourier Transform, El Niño

1 Introduction

Cheng et al. (2015) stated, "Forecasting the evolution of complex systems is noted as one of the 10 grand challenges of modern science." We will show that the DCM can detect the signal(-s) and the trend in complex time series, and can also forecast the time series evolution. Our DCM is based on the LS method, which was originally formulated by Legendre (1805). Gauß (1809) connected this method to the principles of probability and the normal distribution. A few years later, he showed that the LS method gives the best unbiased estimates for the free parameters of linear models, if the zero mean data errors are equal, normally distributed and uncorrelated (Gauß 1821). The extended Gauß-Markov theorem states that the LS method gives the best estimates for the free parameters of linear models, even if the data errors do not pass all the above-mentioned criteria (Wooldridge 2010). The Gauß-Markov theorem is the backbone of our DCM. This time series analysis method performs a massive number of LS fits to find the best model for the data.

Hadamard (1902) defined the three conditions for a well-posed problem. The solution for the problem determines these conditions.

C_1 Existence: A solution exists.
C_2 Uniqueness: The solution is unique.
C_3 Stability: Small changes in the input data cause small changes in the solution.

Stability means that the solution behaves predictably as the input varies. In other words, there exists a continuous mapping from the input space to the solution space. If any of these conditions are not satisfied, the problem is classified as ill-posed.

Ill-posed problems arise in many fields of science (Rudin 1976; Tikhonov and Arsenin 1977; Lavrentiev et al. 1986). Typical examples are partial differential equation solutions (Hadamard 1923), nonlinear parameter estimation (Bard 1974) and regularisation of inverse problems (Engl et al. 1996; Vogel 2002). Ill-posed problems are encountered in time series analysis (Berger 2013; Box et al. 2015) because the models for the data are often non-linear (Tong 1990; Tsay and Chen 2018). The unknown free parameters of the model, namely the frequencies, are in the arguments of trigonometric functions representing the signals. This causes non-linearity because the model partial derivatives contain free parameters of the model.

Our DCM is a frequency-domain parametric time series analysis method. Here, we discuss shortly such methods in the order of increasing model complexity. Numerous methods can be applied if the time series is stationary (constant mean and variance). The Fast Fourier Transform for evenly spaced data (Cooley and Tukey 1965) and the DFT for unevenly spaced data (Lomb 1976; Scargle 1982) are the most widely-used methods. The FFT and DFT assume that the correct model for the time series is a pure sine. The spectral estimating techniques fail if the data are non-stationary. Therefore, trends changing the sample mean and/or variance must be removed before applying spectral estimation (Nerlove 1964). There are numerous methods for detecting one signal superimposed on a trend, especially for evenly spaced data (Cleveland et al. 1990; Shumway and Stoffer 2006; Brockwell and Davis 2009). Kay and Marple (1981) compared many spectral estimating techniques that can detect more than one signal from evenly spaced data. They discussed how the sample window (ΔT) limits the frequency resolution for detecting many signals and causes leakage that shifts power from one spectral peak to another. Exactly correct frequency values may not be detected because the leakage can shift the periodogram peaks. Ghaderpour et al. (2021) presented different techniques for reducing spectral leakage. They also discussed techniques for detecting many signals in unevenly spaced data, if these signals are pure sines. There are more complex modelling problems. For example, the signals are not necessarily pure sines or the sample window ΔT is shorter than the signal period(-s).

Reliable forecasting is the ultimate test for any time series analysis method (Cheng et al. 2015). Forecasting can work for linear and stationary processes, but it is challenging for nonlinear or non-stationary ones. Other forecasting challenges are overfitting and forecast error estimation (Makridakis et al. 1987; Lefrancois 1989; Grushka-Cockayne et al. 2017; Petropoulos et al. 2022).

Our DCM is a frequency-domain parametric time series analysis method. Such methods have their own particular limitations that constrain their applications. Here is our list of those application limitations (AL).

1. Data errors (level of noise) are unknown.
2. Data error information is not utilised.
3. Data must be evenly spaced.
4. Model parameter errors are unknown.
5. Model and forecast errors are unknown.
6. Sample window is shorter than signal period(s).
7. Presence and shape of trend are unknown.
8. Sample window causes leakage.
9. Leakage weakens frequency resolution.
10. Signal shapes are not pure sines.
11. Number of signals is unknown.
12. Correct model alternative is unknown.
13. Signal significances are unknown.
14. Model solution is ill-posed.
15. Complex non-linear model forecasts fail.

This study proceeds through following stages. We present the DCM formulation (Section 2.1). The WD-effect is discussed (Section 2.2). Complex time series are simulated using the DCM model (Sect. 2.3). The DFT and its ALs are presented (Section 2.4). We compare our DCM to the

renowned DFT. The time series analysis of simulated data sets shows that the ALs of DFT do not constrain DCM (Sections 3.1-3.8). Then, we present the identification of the best DCM model for the data (Section 3.9), the DCM significance estimates (Section 3.10), the well-posed computational DCM model solution (Section 3.11) and the DCM forecasts (Section 3.12). Finally, we discuss our results (Section 4).

2 Methods

The observations are $y_i = y(t_i) \pm \sigma(t_i)$, where t_i are the observing times and σ_i are the errors ($i = 1, 2, ..., n$). The sample window is $\Delta T = t_n - t_1$. The mid point is $t_{\rm mid} = t_1 + \Delta T/2$. The mean and the standard deviation of all y_i values are denoted with m and s.

2.1 Discrete Chi-square method (DCM)

Jetsu (2020) introduced the DCM which is a frequency-domain parametric time series analysis method. The primary objective of this current paper is to show that the DCM outperforms all other similar methods.

The DCM model is

$$g(t) = g(t, K_1, K_2, K_3) = h(t) + p(t), \qquad (1)$$

where the integer values K_1, K_2 and K_3 are called the model orders. The notation $g_{K_1,K_2,K_3}(t)$ is used to specify these orders. The DCM model is a sum of two functions. These functions are the periodic function

$$h(t) = h(t, K_1, K_2) = \sum_{i=1}^{K_1} h_i(t, f_i) \qquad (2)$$

$$h_i(t, f_i) = \sum_{j=1}^{K_2} B_{i,j} \cos(2\pi j f_i t) + C_{i,j} \sin(2\pi j f_i t) \qquad (3)$$

and the aperiodic function

$$p(t) = p(t, K_3) = \begin{cases} 0, & \text{if } K_3 = -1 \\ \sum_{k=0}^{K_3} p_k(t), & \text{if } K_3 = 0, 1, 2, ... \end{cases} \qquad (4)$$

where

$$p_k(t) = M_k \left[\frac{2(t - t_{\rm mid})}{\Delta T} \right]^k. \qquad (5)$$

For $k > 1$, the $p_k(t)$ function full range is

$$\begin{cases} 2|M_k|, & \text{if } k \text{ is odd} \\ |M_k|, & \text{if } k \text{ is even.} \end{cases} \qquad (6)$$

The free parameters of $g(t)$ model are

$$\begin{aligned} \boldsymbol{\beta} &= [\beta_1, \beta_2, ..., \beta_\eta] \\ &= [B_{1,1}, C_{1,1}, f_1, ..., B_{K_1,K_2}, C_{K_1,K_2}, f_{K_1}, \\ &\quad M_0, ..., M_{K_3}]. \end{aligned} \qquad (7)$$

The number of free parameters is

$$\eta = K_1 \times (2K_2 + 1) + K_3 + 1. \qquad (8)$$

We divide the free parameters $\boldsymbol{\beta}$ into two groups $\boldsymbol{\beta}_I$ and $\boldsymbol{\beta}_{II}$. The first group are the frequencies

$$\boldsymbol{\beta}_I = [f_1, ..., f_{K_1}]. \qquad (9)$$

Due to this group, all free parameters are not eliminated from all partial derivatives $\partial g/\partial \beta_i$. This makes the $g(t)$ model *non-linear*. If the $\boldsymbol{\beta}_I$ frequencies have constant values, the multipliers $2\pi j f_i$ in Equation 3 become constants. In this case, the model becomes *linear* because the partial derivatives $\partial g/\partial \beta_i$ no longer contain any free parameters. The LS solution for the second group of remaining free parameters

$$\begin{aligned} \boldsymbol{\beta}_{II} = [&B_{1,1}C_{1,1}, ..., B_{K_1,K_2}, C_{K_1,K_2}, \\ &M_0, ..., M_{K_3}] \end{aligned} \qquad (10)$$

becomes *unique*. This solution passes the C_1, C_2 and C_3 conditions of a well-posed problem.

Let us assume that we search for periods between $f_{\rm min}$ and $f_{\rm max}$. The non-linear $g(t)$ model becomes linear, if the tested $\boldsymbol{\beta}_I$ frequencies are fixed to any constant values. The sum $h(t)$ of signals $h_1(t, f_1)$, $h_2(t, f_2)$ ... and $h_{K_1,f_{K_1}}(t)$ does not depend on the order in which these signals are added. For example, the two signal $K_1 = 2$ model symmetry is

$$h(t) = h_1(t, f_1) + h_2(t, f_2) \qquad (11)$$

$$= h_1(t, f_2) + h_2(t, f_1)$$

Both (f_1, f_2) and (f_2, f_1) combinations give the same value for the $g(t)$ model. Since this symmetry applies to any K_1 number of signals, we compute the linear $g(t)$ models only for all tested frequency combinations

$$f_{\max} \geq f_1 > f_2 > ... > f_{K_1} \geq f_{\min}. \quad (12)$$

This frequency space symmetry idea reduces CPU consumption dramatically. For example, were this symmetry not exploited, the four signal $K_1 = 4$ search would give $4! = 24$ exactly the same solutions from four-dimensional tested grid in frequency space. This would cause unnecessary use of CPU in the preliminary testing all possible frequency combinations inside a tesseract (a four-dimensional cube). The search for the best frequency combination from $4! = 24$ different nonlinear iterations would be a pointless exercise (see Equation 22).
The DCM model residuals

$$\epsilon_i = y(t_i) - g(t_i) = y_i - g_i. \quad (13)$$

give the sum of squared residuals

$$R = \sum_{i=1}^{n} \epsilon_i^2, \quad (14)$$

and the Chi-square

$$\chi^2 = \sum_{i=1}^{n} \frac{\epsilon_i^2}{\sigma_i^2}. \quad (15)$$

For every tested $\boldsymbol{\beta}_I = [f_1, f_2, ..., f_{K_1}]$ frequency combination, the LS fit gives the DCM test statistic

$$z = z(f_1, f_2, ..., f_{K_1}) = \sqrt{R/n}, \text{ unknown } \sigma_i \quad (16)$$
$$z = z(f_1, f_2, ..., f_{K_1}) = \sqrt{\chi^2/n}, \text{ known } \sigma_i \quad (17)$$

computed for a linear model. The value of z is unique. The errors can be unknown in Equation 16 (AL1). For known errors, DCM uses this information in Equation 17 (AL2). In the preliminary long search, we test an evenly spaced grid of n_{L} frequencies between $f_{\min}$ and $f_{\max}$. This search gives the best frequency candidates $f_{1,\mathrm{mid}}, ...f_{\mathrm{K_1,mid}}$.

In the final short search, we test a denser grid of n_{S} frequencies within an interval

$$[f_{\mathrm{i,mid}} - a, f_{\mathrm{i,mid}} + a], \quad (18)$$

where $i = 1, .., K_1$, $a = c(f_{\min} - f_{\max})/2$ and $c = 0.05$.

The total number of all tested long and short search frequency combinations is

$$\binom{n_f}{K_1} = \frac{n_f!}{K_1!(n_f - K_1)!}, \quad (19)$$

where $n_f = n_L$ and $n_f = n_S$, respectively. The even or uneven data spacing is irrelevant because the LS fit result for every tested frequency combination does not depend on this spacing (AL3).

The global periodogram minimum

$$z_{\min} = z(f_{1,\mathrm{best}}, f_{2,\mathrm{best}}, ..., f_{\mathrm{K_1,best}}) \quad (20)$$

is at the tested frequencies $f_{1,\mathrm{best}}$, $f_{2,\mathrm{best}}, ...,$ $f_{\mathrm{K_1,best}}$. This tested frequency combination gives the best linear model for the data. The periodogam value z is a scalar, which is computed from K_1 frequency values. It is possible to plot the $K_1 = 2$ two signal periodogram $z(f_1, f_2)$ as a map, where f_1 and f_2 are the coordinates, and $z = z(f_1, f_2)$ is the height. For more than two signals, there is no direct graphical z plot because that requires more than three dimensions. Our solution for this dimensional problem is simple. We plot only the following one-dimensional slices of the full periodogram

$$\begin{aligned}
z_1(f_1) &= z(f_1, f_{2,\mathrm{best}}, ..., f_{\mathrm{K_1,best}}) \quad (21)\\
z_2(f_2) &= z(f_{1,\mathrm{best}}, f_2, f_{3,\mathrm{best}}, ..., f_{\mathrm{K_1,best}})\\
z_3(f_3) &= z(f_{1,\mathrm{best}}, f_{2,\mathrm{best}}, f_3, f_{4,\mathrm{best}}, ..., f_{\mathrm{K_1,best}})\\
z_4(f_4) &= z(f_{1,\mathrm{best}}, f_{2,\mathrm{best}}, f_{3,\mathrm{best}}, f_4, f_{5,\mathrm{best}}, f_{\mathrm{K_1,best}})\\
z_5(f_5) &= z(f_{1,\mathrm{best}}, f_{2,\mathrm{best}}, f_{3,\mathrm{best}}, f_{4,\mathrm{best}}, f_5, f_{\mathrm{K_1,best}})\\
z_6(f_6) &= z(f_{1,\mathrm{best}}, f_{2,\mathrm{best}}, f_{3,\mathrm{best}}, f_{4,\mathrm{best}}, f_{5,\mathrm{best}}, f_6).
\end{aligned}$$

In the above-mentioned $K_1 = 2$ map, the slice $z_1(f_1)$ would represent the height z at the location $(f_1, f_{2,\mathrm{best}})$ when moving along the straight constant line $f_2 = f_{2,\mathrm{best}}$ that crosses the global minimum $z_{\min}$ (Equation 20) at the coordinate point $(f_{1,\mathrm{best}}, f_{2,\mathrm{best}})$. These one-dimensional periodogram slices (Equation 21) allow us "to see inside" the multi-dimensional structure of the

DCM test statistic z (Equations 16 and 17). This kind of visualisation is important in time series analysis (Su and Wu 2024).

The short search gives the best $f_{1,\mathrm{best}}, ..., f_{\mathrm{K}_1,\mathrm{best}}$ frequencies for the data. These frequencies are the *unique* initial values for the first group of free parameters $\boldsymbol{\beta}_{\mathrm{I,initial}} = [f_{1,\mathrm{best}}, ..., f_{\mathrm{K}_1,\mathrm{best}}]$ (Equation 9). The linear model for these constant $[f_{1,\mathrm{best}}, ..., f_{\mathrm{K}_1,\mathrm{best}}]$ frequencies gives the *unique* initial values for the second group $\boldsymbol{\beta}_{\mathrm{II,initial}}$ of free parameters (Equation 10). The non-linear iteration

$$\boldsymbol{\beta}_{\mathrm{initial}} = [\boldsymbol{\beta}_{\mathrm{I,initial}}, \boldsymbol{\beta}_{\mathrm{II,initial}}] \rightarrow \boldsymbol{\beta}_{\mathrm{final}} \tag{22}$$

gives the final free parameter values $\boldsymbol{\beta}_{\mathrm{final}}$.

Furlan and Mortarino (2020) emphasise that the analytical error estimates for the non-linear model free parameters can be tricky. They compare different analytical free parameter error estimating methods by using the computational statistical bootstrap technique (Efron and Tibshirani 1986). They conclude that the analytical error estimates become less reliable if the number of free parameters increases. For our non-linear DCM model, we determine the i:th signal parameters

$$P_i = 1/f_i = \text{Period} \tag{23}$$
$$A_i = \text{Peak to peak amplitude} \tag{24}$$
$$t_{\mathrm{i,min},1} = \text{Deeper primary minimum} \tag{25}$$
$$t_{\mathrm{i,min},2} = \text{Secondary minimum (if present)} \tag{26}$$
$$t_{\mathrm{i,max},1} = \text{Higher primary maximum} \tag{27}$$
$$t_{\mathrm{i,max},2} = \text{Secondary maximum (if present)} \tag{28}$$

and the trend parameters

$$M_k = \text{Polynomial coefficients.} \tag{29}$$

Of these model parameters, all $P_i = 1/f_i$ and all M_k estimates are among the free parameters $\boldsymbol{\beta}$. However, the A_i, $t_{\mathrm{i,min},1}$, $t_{\mathrm{i,min},2}$, $t_{\mathrm{i,max},1}$ and $t_{\mathrm{i,max},2}$ model parameters depend on $\boldsymbol{\beta}$ values. The analytical solutions for these model parameters are simple for pure sines ($K_2 = 1$), but become quite complicated for double waves ($K_2 = 2$). Clearly, the analytical solution for the errors of DCM model parameters would be tedious. The analytical solution for the model error

$$g(t) \pm \sigma_{g(t)} \tag{30}$$

would be practically impossible because the number of DCM model free parameters is large, especially if the data contains many signals.

We use the computational statistical bootstrap technique (Efron and Tibshirani 1986) to solve these analytical problems. In our bootstrap, the tested frequencies are the same as in the short search. We select a random sample $\boldsymbol{\epsilon}^*$ from the residuals $\boldsymbol{\epsilon}$ of the DCM model (Equation 13). Any ϵ_i value can be chosen to the $\boldsymbol{\epsilon}^*$ sample as many times as the random selection happens to favour it. We create $J = 1, 2, ..., n_{\mathrm{B}}$ residual random samples $\boldsymbol{\epsilon}^*_J$. Every $\boldsymbol{\epsilon}^*_J$ sample gives one *artificial* bootstrap data set

$$\boldsymbol{y}^*_J = \boldsymbol{g} + \boldsymbol{\epsilon}^*_J. \tag{31}$$

Each $\boldsymbol{y}^*_J$ sample gives *one* free parameter estimate $\boldsymbol{\beta}_J$. This bootstrap procedure gives n_{B} free parameter estimates $\boldsymbol{\beta}_J$. The standard deviations for all n_{B} estimates for $P_i = 1/f_i$ and all M_k give the errors of these model parameters. We use each $\boldsymbol{\beta}_J$ to compute the model for a dense grid of time points. This gives us n_{B} *numerical* A_i $t_{\mathrm{i,min},1}$, $t_{\mathrm{i,min},2}$, $t_{\mathrm{i,max},1}$ and $t_{\mathrm{i,max},2}$ estimates. The standard deviation of those n_{B} estimates is the error of these model parameters. The $\boldsymbol{\beta}_J$ values give n_{B} estimates for $g(t)$ at any time t. The standard deviation of these $g(t)$ estimates gives the error $\sigma_{g(t)}$ of Equation 30. Note that these $\boldsymbol{\beta}_J$ values can also be used to compute the errors for $h(t), h_i(t), p(t)$ and $p_k(t)$. Our computational statistical bootstrap approach gives not only the errors estimates for the DCM model parameters (AL4), but also the model error inside ΔT and the forecast error outside ΔT (AL5: Equation 30).

There are, of course, totally wrong DCM models for the data. For example, DCM can be forced search for too few or too many K_1 signals, or the selected $p(t)$ trend order K_3 can be wrong, as shown in Figures 5-10 by Jetsu (2020). Such DCM models are unstable and we denote them with "UM", like in Jetsu[1] (2025). These unstable models have three signatures

Intersecting frequencies
Dispersing amplitudes
Leaking periods

[1] Jetsu L. "Do planets cause the sunspot cycle?", submitted to Scientific Reports. August 13th, 2025.

Intersecting frequencies occur when the signal frequencies in the data are very close to each other. We give the following example of how this instability can arise in the two signal model. If the frequency f_1 approaches the frequency f_2, both $h_1(t)$ and $h_2(t)$ signals become essentially one and the same signal. The LS fit fails because it makes no sense to model the same signal twice.

Dispersing amplitudes instability can occur, if the two signal frequencies are too close to each other. The LS fit finds a model, where two high amplitude signals nearly cancel out each other. The low amplitude signal, the sum of these two high amplitude signals, fits to the data.

There are DCM models where the detected frequency f is outside the tested frequency interval between $f_{\min}$ and $f_{\max}$. This leaking periods instability may indicate that the chosen tested period range is wrong.

The DCM model (Equation 1) is more sophisticated than the models of our former time series analysis methods, the Three Stage Period Search (Jetsu and Pelt 1999, TSPA, Equation 1) and the Continuous Period Search (Lehtinen et al. 2011, CPS, Equation 3). TSPA and CPS can detect only one signal ($K_1 = 1$) from stationary time series ($K_3 = 0$). TSPA can detect pure sine and double wave signals ($K_2 = 1$ or 2). The extension of TSPA, the CPS method, tests three alternatives: $K_2 = 1$ or 2, or no signal at all. The DCM model is more sophisticated because it can have any arbitrary K_1, K_2 and K_3 combination. The DCM sum $g(t) = h(t) + p(t)$ of arbitrary periodic and aperiodic functions represents a universal model because the innumerable K_1, K_2 and K_3 combinations allow unlimited complexity. Even more complex signal shape (K_2) combinations could be used. For example, these shapes could be $K_{2,i=1} = 3$ (3rd harmonic), $K_{2,i=2} = 1$ (pure sine) and $K_{2,i=3} = 2$ (double wave), where if $K_{2,i}$ denotes the shape of i:th signal.

2.2 WD-effect: Spearhead of DCM

In this section, we discuss what causes the WD-effect defined in our abstract:

> *"For any sample window ΔT, DCM inevitably detects the correct $p(t)$ trend and $h(t)$ signal(-s) when the sample size n and/or data accuracy σ increase."*

The consequences of this effect are also discussed.

The t_i, y_i and σ_i data are inside the rectangle $\Delta T \times \Delta y = (t_n - t_1) \times [\max(y_i + \sigma_i) - \min(y_i - \sigma_i)]$. The LS fit results do not depend on ΔT. If the measuring time intervals δt_i for each observation y_i fulfil $\delta t_i \ll \Delta T$, the R and χ^2 values obtained from *all* LS (Equation 19) depend *only* on y_i changes in ΔY direction, but not on the t_i changes in ΔT direction. For fixed y_i and σ_i, the residuals ϵ_i determine the R and χ^2 values. These residuals measure *only* the Δy direction. Hence, the periodogram z values (Equation 16 or 17) do not depend on ΔT. If ΔT decreases, the R and χ^2 estimates can measure the g_i model details inside the $\Delta T \times \Delta y$ rectangle only if n increases and/or σ decreases. Better data reveal these model details.

Plenty of concrete examples will confirm that the WD-effect is real (Sections 3.1-3.7). Our DCM can detect signals(-s) when the sample window ΔT is shorter than the period P value(-s). The DCM surpasses AL6. The data window ΔT length is irrelevant. This means that the DCM can model an infinitesimal time series, as well as forecast its future and past, if the sample size (n) and/or the data accuracy (σ) are sufficient. That is a revolutionary achievement.

2.3 Simulated DCM model time series samples

We use seven different $g(t)$ models to simulate 21 artificial complex time series (Sections 3.1-3.7). The sample window of all simulated time series is $\Delta T = 1$. The $n^\star$ simulated time points $t_i^\star$ are drawn from a random uniform distribution $U(0, \Delta T, n^\star)$. The first and last time point values are then modified to $t_1^\star = 0$ and $t_n^\star = \Delta T$. Hence, the distance between independent frequencies (Kay and Marple 1981) is always

$$f_0 = 1/\Delta T = 1. \tag{32}$$

The $n^\star$ residuals $\epsilon^\star(t_i^\star)$ of simulated model are drawn from a random normal distribution $N(0, \sigma, n^\star)$, where σ is the accuracy of simulated data. The simulated data are

$$\begin{aligned} y^\star(t_i^\star) &= g(t_i^\star) + \epsilon^\star(t_i^\star) \\ &= h(t_i^\star) + p(t_i^\star) + \epsilon^\star(t_i^\star). \end{aligned}$$

Table 1 Model 1: DCM analysis between $P_{\min} = 0.63$ and $P_{\max} = 5.70$. (1) Simulated P_1, A_1, $t_{1,\min,1}$, $t_{1,\max,1}$ and M_0 values. (2-4) Detected values for different n and SN combinations. Two lowest lines specify electronic data files and DCM analysis control files.

(1) Model 1	(2) $n = 50$ SN = 10	(3) $n = 50$ SN = 50	(4) $n = 100$ SN = 100
$P_1 = 1.9$	1.58 ± 0.21	1.822 ± 0.042	1.863 ± 0.017
$A_1 = 2.0$	1.64 ± 0.33	1.894 ± 0.072	1.941 ± 0.025
$t_{1,\min,1} = 1.35$	1.20 ± 0.11	1.312 ± 0.021	1.3321 ± 0.0088
$t_{1,\max,1} = 0.40$	0.4098 ± 0.0058	0.4007 ± 0.0013	0.40056 ± 0.00078
$M_0 = 1.0$	1.20 ± 0.18	1.061 ± 0.036	1.030 ± 0.014
Data file	`Model1n50SN10.dat`	`Model1n50SN50.dat`	`Model1n100SN100.dat`
Control file	`dcmModel1n50SN10.dat`	`dcmModel1n50SN50.dat`	`dcmModel1n100SN100.dat`

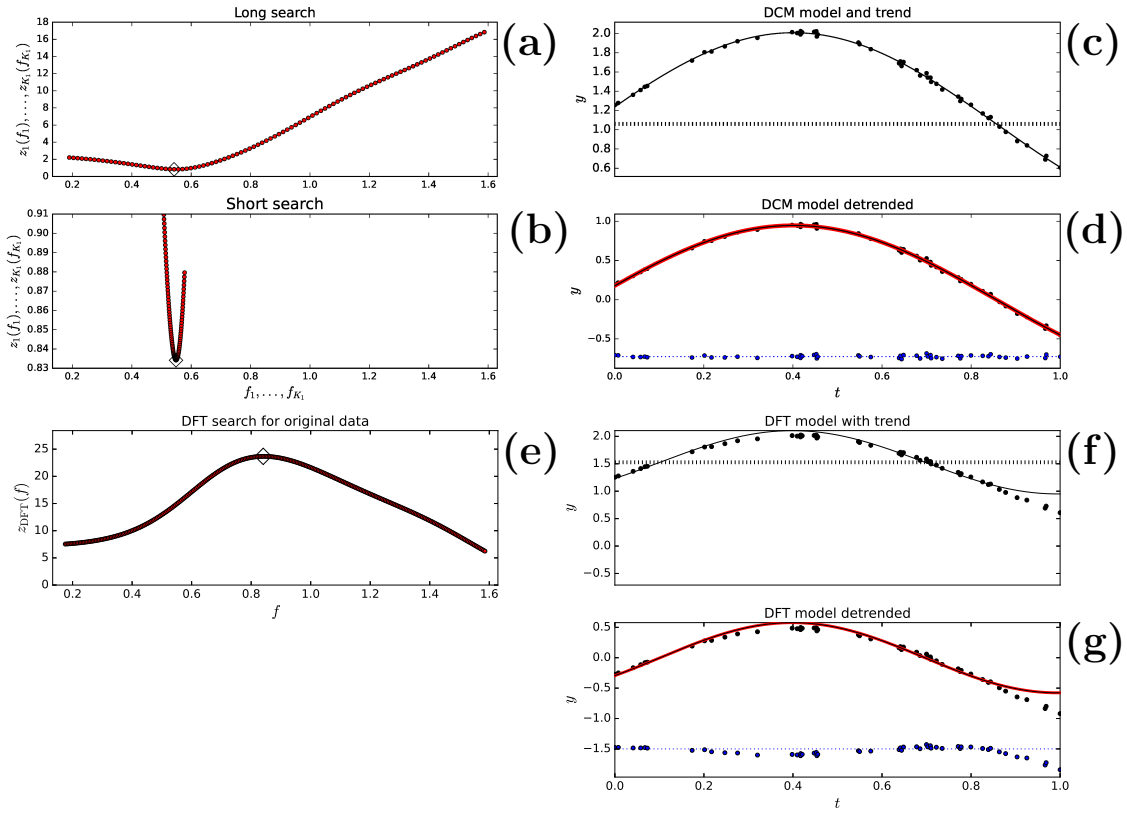

Fig. 1 Model 1 (Table 1: $n = 50$, SN = 50 simulation). (a) DCM long search periodogram $z_1(f_1)$ gives best period at 1.843 (diamond). (b) DCM short search periodogram $z_1(f_1)$ gives best period at 1.822 (diamond). (c) DCM model $g(t)$ (black continuous line), DCM trend $p(t)$ (black dashed line) and data y_i (black dots). (d) DCM model detrended $g(t) - p(t)$ (black continuous line), DCM signal $h_1(t)$ (red thick continuous line), detrended data $y(t_i) - p(t_i)$ (black dots) and DCM model residuals $y(t_i) - g(t_1)$ (blue dots) offset to -0.65 level (blue dotted line). (e) DFT periodogram $z_{\rm DFT}(f)$ gives best period at 1.190 (Diamond). (f) DFT model $g_{\rm DFT}(t)$ (black continuous line), DFT trend $p_{\rm DFT}(t)$ (black dashed line) and data y_i (black dots). (g) DFT model detrended $g_{\rm DFT}(t) - p_{\rm DFT}(t)$ (black continuous line), DFT pure sine $s_{\rm DFT}(t)$ (red thick continuous line), detrended data $y(t_i) - p_{\rm DFT}(t_i)$ (black dots) and DFT model residuals (blue dots) offset to -1.5 level (blue dotted line).

The peak to peak amplitudes of all simulated signals is $A = 2$. Our definition for the signal to noise ratio is

$$\mathrm{SN} = (A/2)/\sigma. \quad (33)$$

2.4 Discrete Fourier Transform (DFT)

The DFT is one of the most prevalent frequency-domain parametric time series analysis methods for unevenly spaced time series. It searches for the best pure sine model for the data. The equivalent DCM model has the orders $K_1 = 1, K_2 = 1$ and $K_3 = 0$. Any time series analysis method must be remarkable if it performs better than the distinguished DFT. Therefore, we will compare how DCM and DFT perform in the analyses of simulated time series (Sections 3.1-3.7). We search for signals in these simulated time series by using the frequently applied DFT version formulated by Horne and Baliunas (1986)[2]. Our notation for the DFT test statistic is $z_{\mathrm{PDF}}(f)$ (Horne and Baliunas 1986, Equation 1). The notations for the DFT model are

$$g_{\mathrm{DFT}}(t) = s_{\mathrm{DFT}}(t) + p_{\mathrm{DFT}}(t), \quad (34)$$

where $s_{\mathrm{DFT}}(t)$ is the sum of pure sine signals and $p_{\mathrm{DFT}}(t)$ is the trend. The pure sine signals for the y_i data and the ϵ_i residuals are denoted with $s_{y,\mathrm{DFT}}(t)$ and $s_{\epsilon,\mathrm{DFT}}(t)$, respectively. Our DFT analyses of simulated time series may fail due to AL6-AL10.

The time series is "too short" (Kay and Marple 1981, AL6)

$$P > \Delta T. \quad (35)$$

The time series is non-stationary because it contains "a trend" (Nerlove 1964, AL7)

$$p_{\mathrm{DFT}}(t) \neq 0. \quad (36)$$

Due to the leakage caused by the sample window (AL8), the signal frequencies are "too close" (Kay and Marple 1981, AL9).

$$|f_1 - f_2| < f_0 = \Delta T^{-1}. \quad (37)$$

[2]This paper had over 2600 citations in December 2025

The signals are not "pure sines" (Bretthorst 1988, AL10)

$$K_2 \neq 1. \quad (38)$$

Many parametric frequency-domain time series analysis methods, like the DFT, can be applied only to stationary data. Trends changing the time series mean or variance must be removed before applying these methods (Nerlove 1964). The DFT can detect only one period at the time. Such frequency-domain parametric time series analysis methods are hereafter called "one-dimensional". Our DFT analysis of simulated data proceeds through two stages. First, the $p_i = p_{\mathrm{DFT}}(t_i)$ trend is removed from the simulated data $y_i = y(t_i)$. Then, the iterative pre-whitening technique described by Ghaderpour et al. (2021) is applied to search for the pure sine signals. We search for the first signal by applying the DFT to the detrended $y_i - p_i$ simulated data. In the second signal search, the DTF is applied to the original model residuals $\epsilon_i = (y_i - p_i) - g_{\mathrm{y,DFT}}(t_i)$. At any stage, this combination of detrending and iterative pre-whitening analysis may fail and corrupt the results obtained at the next stages. For example, all DFT analyses of simulated time series will fail already at the detrending stage (Sections 3.1-3.7). Our DCM has no corrupting separate stages because the trend and the signal(-s) are detected (modelled) simultaneously.

3 Results

We apply DCM and DFT to seven different simulated time series (Sections 3.1 - 3.7). As we proceed, these time series become increasingly complex. DCM analysis succeeds for all time series. DFT analysis fails for every time series.

3.1 Model 1

Our first time series simulation model is the one signal model

$$g(t) = (A_1/2)\cos\left[\frac{2\pi(t - t_{1,\max,1})}{P_1}\right] + M_0. \quad (39)$$

We give the P_1, A_1, $t_{1,\max,1}$ and M_0 values in Table 1 (Column 1). This sample is "too short" (Equation 35) because $P_1 = 1.9\Delta T$. The constant mean level M_0 is unknown (Equation 36).

We perform the DCM and DFT time series analysis between $P_{\min} = P_1/3 = 0.63$ and $P_{\max} = 3P_1 = 5.70$.

Model 1 is a DCM model, where $K_1 = 1$, $K_2 = 1$, $K_3 = 0$, $B_{1,1} = (A_1/2)\cos(2\pi f_1 t_{1,\max,1})$ and $C_{1,1} = (A_1/2)\sin(2\pi f_1 t_{1,\max,1})$. We give the DCM analysis results for three samples having different n and SN combinations (Table 1: Columns 2-4). For each sample, this table specifies the electronic data file and the electronic DCM control file.[3] The detected P_1, A_1, $t_{1,\max,1}$ and M_0 values are correct and accurate even for the combination $n = 50$ and SN = 50. Regardless of $\Delta T < P_1$, these model parameter values become more accurate and converge to the correct simulated values when n and SN increase. This confirms the WD-effect.

A graphical presentation of DCM analysis results is shown for Model 1 simulated time series, where $n = 50$ and SN = 50 (Figures 1a-d). The DCM long search $z_1(f_1)$ periodogram minimum is at $P_1 = 1.843$ (Figure 1a: diamond). The DCM short search periodogram $z_1(f_1)$ gives the best period at $P_1 = 1.822$ (Figure 1b: diamond). The continuous black line denoting the DCM model $g(t)$ fits perfectly to the black dots denoting the data y_i (Figure 1c). The mean $p(t) = M_0 = 1.061 \pm 0.036$ is correct (Figure 1c: dashed black line). The detrended model $g(t) - p(t)$ (black continuous line), the detrended data $y(t_i) - p(t_i)$ (black dots) and the pure sine signal $h_1(t)$ (red thick continuous line) are shown in Figure 1d. Note that the thick continuous red line stays under the thin continuous black line because $h_1(t) = g(t) - p(t)$. The DCM residuals (blue dots) are offset to the level of -0.65 (blue dotted line). These residuals are stable and display no trends.

The DFT detects the wrong period $P_1 = 1.190$ (Figure 1e: Diamond). The DFT mean level estimate $M_0 = 1.527$ is also wrong (Figure 1f: Dashed black line). The black dots denoting the data y_i deviate from the continuous black line denoting the DFT model $g_{\mathrm{DFT}}(t)$. The detrended model $g_{\mathrm{DFT}}(t) - p_{\mathrm{DFT}}(t)$ (continuous black line), the detrended data $y(t_i) - p_{\mathrm{DFT}}(t_i)$ (black dots) and the signal $s_{\mathrm{y,DFT}}(t)$ (continuous thick red line) are shown in Figure 1g. The thin black line covers the thick red line because $s_{\mathrm{y,DFT}}(t) = g_{\mathrm{DFT}}(t) - p_{\mathrm{DFT}}(t)$. The DFT residuals (blue dots) offset to the level of -1.5 (blue dotted line) display obvious trends, especially at the end of analysed sample.

For the simulated time series of Model 1, the DCM analysis succeeds, but the DFT analysis fails.

3.2 Model 2

Our next one signal time series simulation model is

$$g(t) = (A_1/2)\cos\left[\frac{2\pi(t - t_{1,\max,1})}{P_1}\right] + M_0 + M_1 T + M_2 T^2 \quad (40)$$

where $T = [2(t - t_{\mathrm{mid}})]/\Delta T$. We give the P_1, A_1, $t_{1,\max,1}$, M_0, M_1 and M_2 values in Table 2. As a DCM model, the orders of Model 2 are $K_1 = 1$, $K_2 = 1$ and $K_3 = 2$. The simulated time series is "too short" because the period P_1 is $1.9 \times \Delta T$ (Equation 35). The parabolic trend $p(t)$ is unknown (Equation 36). Again, we use the DCM and DFT time series analysis methods to search for periods between $P_{\min} = P_1/3 = 0.63$ and $P_{\max} = 3P_1 = 4.70$.

The DCM analysis results are given in Table 2. These results are not very accurate for the $n = 100$ and SN = 100 combination, but they definitely improve for larger n and SN values. The WD-effect ensures that the detected values converge to the correct simulated model parameter values. The short sample window, $\Delta T < P_1$, does not mislead the DCM analysis.

For Model 2, the DCM analysis results are illustrated for the $n = 10\,000$ and SN = 100 combination (Figures 2a-d). The DCM long search $z_1(f_1)$ periodogram minimum is at $P_1 = 1.843$ (Figure 2a: diamond). The DCM short search gives the value $P_1 = 1.852$ (Figure 2b: diamond). The DCM model $g(t)$ is so good that it's continuous black line is totally covered by the black dots representing the y_i data (Figure 2c). Therefore, the colour of this $g(t)$ line has been changed from black to white. The results for the parabolic trend coefficients $M_0 = 1.079 \pm 0.096$, $M_1 = 0.229 \pm 0.026$ and $M_2 = 0.451 \pm 0,064$ of the dashed black $p(t)$ line are correct. In Figure 1d, the white continuous line shows the detrended model $g(t) - p(t)$. The black dots show the detrended data $y(t_i) - p(t_i)$ and the red thick continuous line

[3] We publish all our data files and all DCM control files in https://zenodo.org/uploads/17018676

Table 2 Model 2. DCM analysis between $P_{\min} = 0.63$ and $P_{\max} = 4.70$. Notations as in Table 1.

(1) Model 2	(2) $n = 1\,000$ SN = 100	(3) $n = 10\,000$ SN = 100	(4) $n = 10\,000$ SN = 200
$P_1 = 1.9$	1.98 ± 0.34	1.852 ± 0.055	1.933 ± 0.040
$A_1 = 2.0$	2.4 ± 2.8	1.84 ± 0.20	2.13 ± 0.16
$t_{1,\min,1} = 1.35$	1.39 ± 0.17	1.326 ± 0.027	1.367 ± 0.020
$t_{1,\max,1} = 0.40$	0.4015 ± 0.0020	0.40007 ± 0.00072	0.40017 ± 0.00075
$M_0 = 1.0$	0.8 ± 1.4	1.079 ± 0.096	0.937 ± 0.080
$M_1 = 0.25$	0.29 ± 0.21	0.229 ± 0.026	0.266 ± 0.020
$M_2 = 0.50$	0.62 ± 0.52	0.451 ± 0.064	0.540 ± 0.049
Data file	`Model2n1000SN100.dat`	`Model2n10000SN100.dat`	`Model2n10000SN200.dat`
Control file	`dcmModel2n1000SN100.dat`	`dcmModel2n10000SN100.dat`	`dcmModel2n10000SN200.dat`

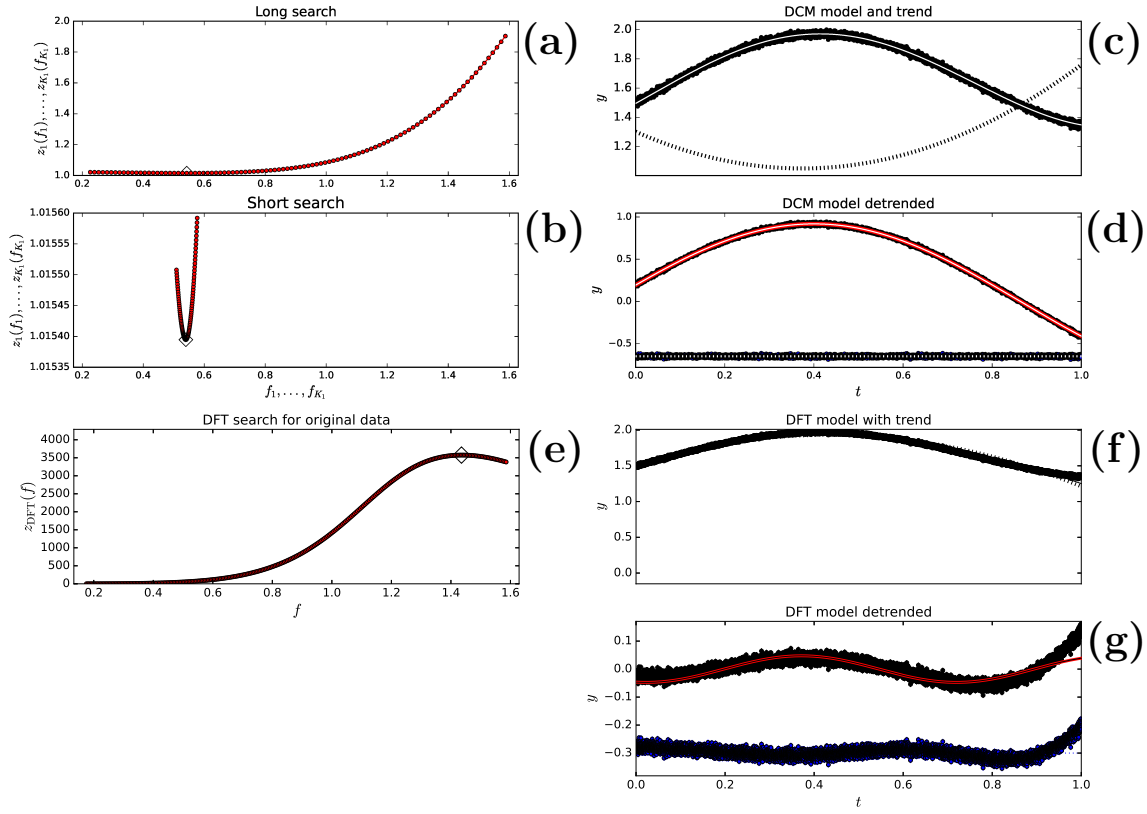

Fig. 2 Model 2 (Table 2: $n = 10\,000$, SN = 100 simulation). (c) Colour of $g(t)$ line has been changed from black to white. (d) Colour of $g(t)$ line has been changed from black to white. Colour of offset level -0.3 dotted line has been changed from blue to white. Locations of best periods (diamonds) are explained in text (Section 3.2). Otherwise, notations are as in Figure 1.

shows the pure sine signal $h_1(t)$. Note that the red thick line is under the white thin line because $h_1(t) = g(t) - p(t)$. The DCM residuals (blue dots) are offset to the level of -0.65. The colour of dotted line, which denotes this offset level, has been changed from blue to white. The distribution of these DCM model residuals is stable, as expected for a random normal distribution.

The wrong period $P_1 = 0.697$ is detected by DFT (Figure 2e: Diamond). The DFT estimates for the trend $p(t)$ coefficients, $M_0 = 1.92$, $M_1 = -0.15$ and $M_2 = -0.56$ are also wrong (Figure

Table 3 Model 3. DCM analysis between $P_{\min} = 0.053$ and $P_{\max} = 0.480$. Notations as in Table 1.

(1) Model 3	(2) $n = 50$ SN = 10	(3) $n = 50$ SN = 50	(4) $n = 100$ SN = 100
$P_1 = 0.16$	0.15898 ± 0.00068	0.15989 ± 0.00014	0.159950 ± 0.000077
$A_1 = 2.0$	1.85 ± 0.10	1.987 ± 0.036	1.994 ± 0.015
$t_{1,\min,1} = 0.11$	0.1122 ± 0.0020	0.11080 ± 0.00088	0.11004 ± 0.00030
$t_{1,\max,1} = 0.03$	0.0328 ± 0.0022	0.03086 ± 0.00094	0.03007 ± 0.00033
$P_2 = 0.17$	0.17025 ± 0.00066	0.17018 ± 0.00024	0.16996 ± 0.00080
$A_2 = 2.0$	2.001 ± 0.090	1.962 ± 0.046	2.002 ± 0.012
$t_{2,\min,1} = 0.135$	0.1312 ± 0.0022	0.13461 ± 0.00087	0.13495 ± 0.00031
$t_{2,\max,1} = 0.05$	0.0461 ± 0.0024	0.04952 ± 0.00099	0.04997 ± 0.000038
$M_0 = 1.0$	0.996 ± 0.011	0.9985 ± 0.0032	0.9995 ± 0.0012
Data file	`Model3n50SN10.dat`	`Model3n50SN50.dat`	`Model3n100SN100.dat`
Control file	`dcmModel3n50SN10.dat`	`dcmModel3n50SN50.dat`	`dcmModel3n100SN100.dat`

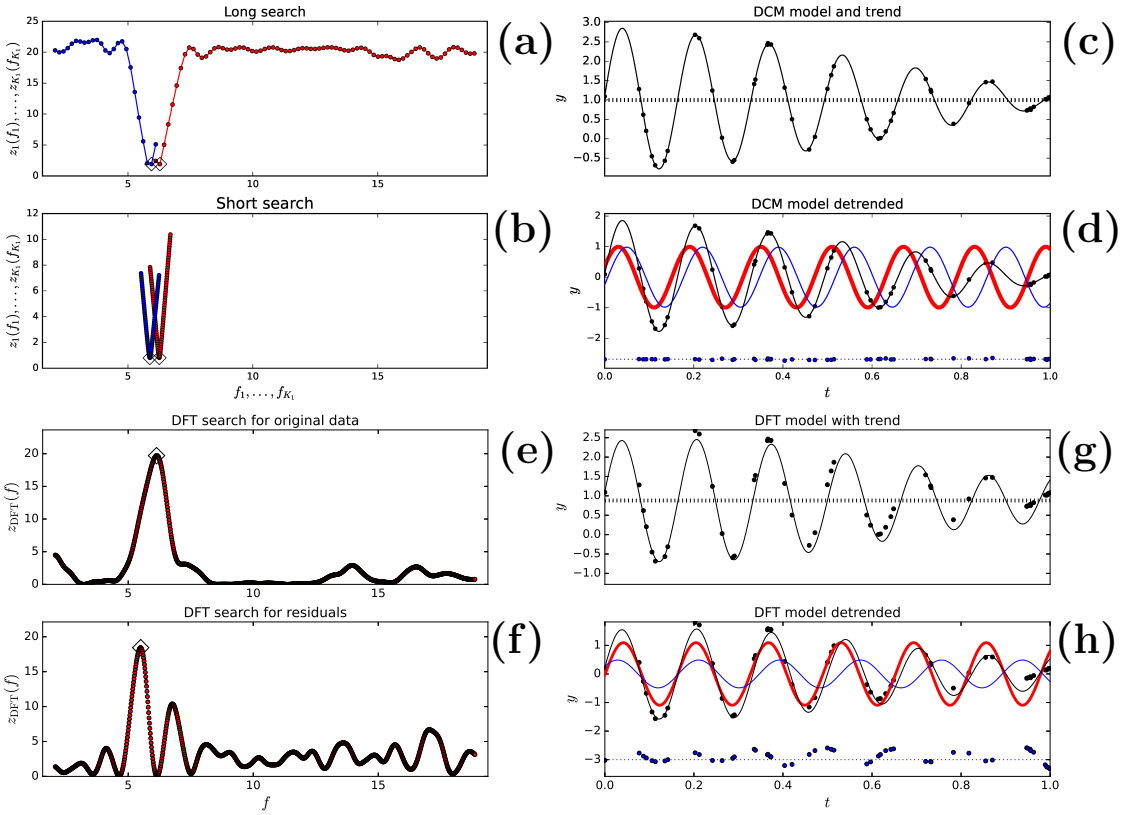

Fig. 3 Model 3 (Table 3: $n = 50$, SN = 50 simulation). (a) DCM long search periodograms $z_1(f_1)$ (red) and $z_2(f)$ (blue) give best periods at 0.160 and 0.168 (diamonds). (b) DCM short search periodograms $z_1(f_1)$ (red) and $z_2(f)$ (blue) give best periods at 0.160 and 0.170 (diamonds). (c) DCM model $g(t)$ (black continuous line), DCM trend $p(t)$ (black dashed line) and data y_i (black dots). (d) DCM model detrended $g(t) - p(t)$ (black continuous line), DCM signal $h_1(t)$ (red thick continuous line), DCM signal $h_2(t)$ (blue continuous thin line), detrended data $y(t_i) - p(t_i)$ (black dots) and DCM model residuals $y(t_i) - g(t_1)$ (blue dots) offset to -3.0 level (blue dotted line) (e) DFT periodogram $z_{\rm DFT}(f)$ for the original data gives best period at 0.163 (diamond). (f) DFT periodogram $z_{\rm DFT}(f)$ for the sine model residuals gives best period at 0.182 (diamond). (g) DFT model $g_{\rm DFT}(t)$ (black continuous line), DFT trend $p_{\rm DFT}(t)$ (black dashed line) and data y_i (black dots). (h) DFT model detrended $g_{\rm DFT}(t) - p_{\rm DFT}(t)$ (black continuous line), DFT pure sine model for original data $s_{y,{\rm DFT}}(t)$ (red thick continuous line), DFT pure sine model for first residuals $s_{\epsilon,{\rm DFT}}(t)$ (blue continuous thin line), detrended data $y(t_i) - p_{\rm DFT}(t_i)$ (black dots) and DFT model residuals (blue dots) offset to -3.0 level (blue dotted line).

Table 4 Model 4. DCM analysis between $P_{\min} = 0.053$ and $P_{\max} = 0.480$. Notations as in Table 1.

(1) Model 4	(2) $n = 50$ SN = 10	(3) $n = 50$ SN = 50	(4) $n = 100$ SN = 100
$P_1 = 0.16$	0.1606 ± 0.0084	0.15982 ± 0.00023	0.16001 ± 0.000058
$A_1 = 2.0$	1.97 ± 0.15	1.958 ± 0.051	2.012 ± 0.014
$t_{1,\min,1} = 0.11$	0.1134 ± 0.0036	0.11059 ± 0.00088	0.10977 ± 0.00033
$t_{1,\max,1} = 0.03$	0.0331 ± 0.0040	0.03068 ± 0.00097	0.02976 ± 0.00035
$P_2 = 0.17$	0.1744 ± 0.0012	0.17009 ± 0.00026	0.169958 ± 0.000092
$A_2 = 2.0$	1.74 ± 0.17	1.982 ± 0.044	2.018 ± 0.017
$t_{2,\min,1} = 0.135$	0.1321 ± 0.0044	0.1342 ± 0.0011	0.13528 ± 0.00042
$t_{2,\max,1} = 0.05$	0.0460 ± 0.0048	0.0492 ± 0.0012	0.05031 ± 0.00034
$M_0 = 1.0$	1.015 ± 0.024	0.9997 ± 0.0030	1.0012 ± 0.0016
$M_1 = 0.25$	0.219 ± 0.023	0.2565 ± 0.0066	0.2495 ± 0.0019
$M_2 = 0.5$	0.485 ± 0.050	0.511 ± 0.011	0.4983 ± 0.0033
Data file	`Model4n50SN10.dat`	`Model4n50SN50.dat`	`Model4n100SN100.dat`
Control file	`dcmModel4n50SN10.dat`	`dcmModel4n50SN50.dat`	`dcmModel4n100SN100.dat`

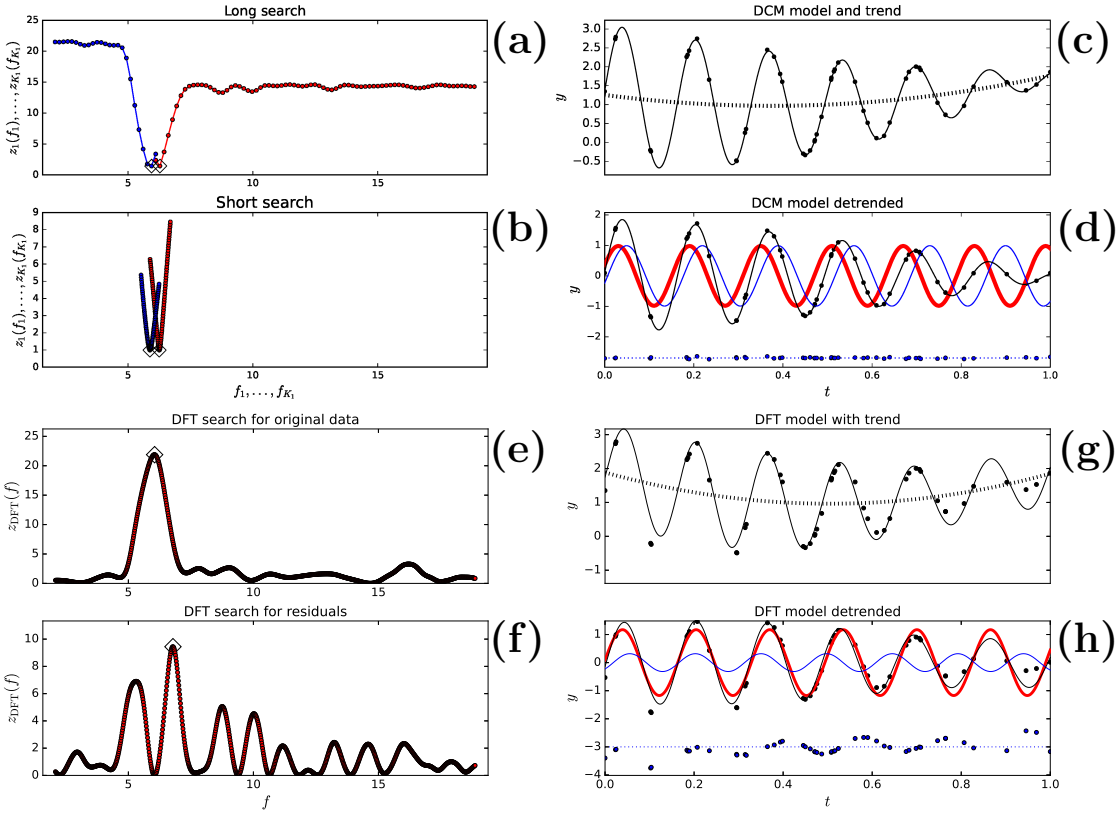

Fig. 4 Model 4 (Table 4: $n = 50$, SN = 50 simulation). Notations as in Figure 3. Best periods (diamonds) are explained in Section 3.4.

2f: dashed black line). The data y_i (black dots) deviate from the DFT model $g_{\mathrm{DFT}}(t)$ (black continuous line), especially in the end of the sample (Figure 2g). For the detrended DFT model, the thin black line covers the thick red line because $s_{\mathrm{y,DFT}}(t) = g_{\mathrm{DFT}}(t) - p_{\mathrm{DFT}}(t)$ (Figure 2g). The $s_{\mathrm{y,DFT}}(t)$ sine curve peak to peak amplitude is far

below the correct $A_1 = 2.0$ value. The DFT residuals (blue dots) are offset to the level of -1.5 (blue dotted line). The trends of these residuals confirm that the DFT analysis fails.

Only DCM (not DFT) succeeds in the analysis of Model 2 simulated time series.

3.3 Model 3

Our third time series simulation model is the two signal model

$$g(t) = (A_1/2)\cos\left[\frac{2\pi(t-t_{1,\max,1})}{P_1}\right] + (A_2/2)\cos\left[\frac{2\pi(t-t_{2,\max,1})}{P_2}\right] + M_0. \quad (41)$$

We give the P_1, P_2, A_1,A_2, $t_{1,\max,1}$, $t_{2,\max,1}$ and M_0 values in Table 3 (Column 1). The DCM orders of Model 3 are $K_1 = 2$, $K_2 = 1$ and $K_3 = 0$. The simulated time series is not "too short" (Equation 35) because $P_1 < P_2 < \Delta T$. The $p(t)$ trend mean level M_0 is unknown (Equation 36). The two frequencies are "too close" because $\Delta f = f_1 - f_2 = f_0/2.72$ (Equation 37). We apply the DCM and DFT time series analyses to search for periods between $P_{\min} = P_1/3 = 0.053$ and $P_{\max} = 3P_1 = 0.480$.

The DCM analysis results for different n and SN combinations are given in Table 3 (Columns 2-4). The results are surprisingly accurate even for the small $n = 50$ sample having a low SN = 10. Due to the WD-effect, the detected model parameter values converge to the correct simulated values when n and SN increase.

The DCM results for Model 3 are illustrated in Figures 1a-d ($n = 50$ and SN = 50 combination). The DCM long search $z_1(f_1)$ periodogram (red) and $z_2(f)$ periodogram (blue) minima are at the best periods $P_1 = 0.159$ and $P_2 = 0.168$ (Figure 3a: diamonds). The DCM short search gives $P_1 = 0.160$ and $P_2 = 0.170$ (Figure 3b: diamonds). The black continuous DCM model $g(t)$ curve crosses trough the black dots of y_i data (Figure 1c). The result $M_0 = 0.9985 \pm 0.0032$ for the dashed black $p(t)$ trend line is correct. Our Figure 3d shows the detrended DCM model $g(t) - p(t)$ (black continuous line), the detrended data $y(t_i) - p(t_i)$ (black dots), the pure sine signal $h_1(t)$ (red thick continuous line) and the pure sine signal $h_2(t)$ (blue thin continuous line). The DCM residuals (blue dots) are offset to the level of -3.0 (blue dotted line). These residuals are small and their level is stable.

The DFT detects the wrong period $P_1 = 1.163$ for the original data (Figure 3e: diamond). This is an expected result because the detected period should be close to $(P_1 + P_2)/2$ when the peak to peak amplitudes of the simulated data, $A_1 = A_2$, are equal (Jetsu 2025). The two DFT periodogram $z_{\rm DFT}(f)$ peaks at frequencies $1/P_1$ and $1/P_2$ overlap and merge into one peak. The period $P_2 = 0.182$ detected for the residuals is also wrong (Figure 3f: diamond). The DFT trend $p_{\rm PDF}(t)$ estimate $M_0 = 0.880$ fails (Figure 3f: dashed black line). The black dots y_i show minor deviations from the continuous black DFT model $g_{\rm DFT}(t)$ line (Figure 3g). The detrended model $g_{\rm DFT}(t) - p_{\rm DFT}(t)$ (continuous black line), the detrended data $y(t_i) - p_{\rm DFT}(t_i)$ (black dots), the pure sine signal $s_{\rm y,DFT}(t)$ for the data (continuous thick red line) and the pure sine signal $s_{\epsilon,\rm DFT}(t)$ for the residuals (continuous blue thin line) are shown in Figure 3h. Note that the $s_{\rm y,DFT}(t)$ and $s_{\epsilon,\rm DFT}(t)$ signal amplitudes are far from equal. The DFT residuals (blue dots) offset to the level of -3.0 (blue dotted line) are not stable.

The DCM analysis of Model 3 simulated time series succeeds, but the DFT analysis does not.

3.4 Model 4

The next time series simulation model is

$$g(t) = (A_1/2)\cos\left[\frac{2\pi(t-t_{1,\max,1})}{P_1}\right] + (A_2/2)\cos\left[\frac{2\pi(t-t_{2,\max,1})}{P_2}\right] + M_0 + M_1 T + M_2 T^2, \quad (42)$$

where $T = [2(t - t_{\rm mid})]/\Delta T$. In this model, two signals are superimposed on an unknown parabolic trend. The P_1, P_2, A_1,A_2, $t_{1,\max,1}$, $t_{2,\max,1}$, M_0, M_1 and M_2 values are given in Table 4. This simulated sample is not "too short" (Equation 35). The polynomial trend $p(t)$ is unknown (Equation 36). The $\Delta f = f_1 - f_2 = f_0/2.72$ difference means that the frequencies are "too close" (Equation 37). The DCM and DFT time series analysis methods are used to search for periods between $P_{\min} = P_1/3 = 0.053$ and $P_{\max} = 3P_1 = 0.480$.

Model 4 is a DCM model having orders $K_1 = 2$, $K_2 = 1$ and $K_3 = 2$. Our DCM analysis results

for different n and SN combinations are given in Table 4 (Columns 2-4). The DCM detects the correct P_1, P_2, A_1, A_2, $t_{\max,1}$, $t_{\max,2}$, M_0, M_1 and M_2 values even for the lowest $n = 50$ and SN = 10 combination. As the simulated data n and SN values increase, the WD-effect ensures that all detected values converge to the correct simulated model parameter values.

Our Figures 4a-d illustrate the DCM analysis results for one sample of simulated time series (Model 4: $n = 50$ and SN = 50). The DCM long search best periods are at $P_1 = 0.159$ and $P_2 = 0.168$ (Figure 4a: diamonds). The DCM short search values are $P_1 = 0.160$ and $P_2 = 0.170$ (Figure 4b: diamonds). The continuous DCM model $g(t)$ black line crosses through all black dots representing the y_i data (Figure 4c). The DCM detects the correct polynomial trend $p(t)$ coefficients $M_0 = 0.9997 \pm 0.0030$, $M_1 = 0.2565 \pm 0.0066$ and $M_2 = 0.511 \pm 0.011$ (Figure 4c: dashed black line). The detrended DCM model $g(t) - p(t)$ (black continuous line), the detrended data $y(t_i) - p(t_i)$ (black dots), the pure sine signal $h_1(t)$ (red thick continuous line) and the pure sine signal $h_2(t)$ (blue thin continuous line) are shown in Figure 4d. The DCM residuals (blue dots) offset to the level of -3.0 show no trends and are extremely stable.

Since the peak to peak amplitudes of the simulated data signals are equal, $A_1 = A_2$, the expected result for the DFT analysis of original data is $(P_1 + P_2)/2$, where P_1 and P_2 are the simulated signal periods (Jetsu 2025). The DFT detects this expected wrong period $P_1 = 1.165$ for the original data (Figure 4e: diamond). A wrong period $P_2 = 0.147$ is also detected for the residuals (Figure 4f: diamond). The DFT estimate $M_0 = 0.962$ for the $p(t)$ trend is close to the correct value $M_0 = 1$, but the $M_1 = -0.017$ and $M_2 = 0.903$ estimates are wrong (Figure 4g: dashed black line). The continuous black DFT model $g_{\rm DFT}(t)$ line deviates from the black dots of data y_i, especially in the beginning and end of the sample. In our Figure 4h, the black dots are the detrended data $y(t_i) - p_{\rm DFT}(t_i)$ and the continuous black line is the detrended DFT model $g_{\rm DFT}(t) - p_{\rm DFT}(t)$. The continuous thick red line is the pure sine signal $s_{\rm y,DFT}(t)$ for the original data and the continuous thinner blue line is the pure sine signal $s_{\epsilon,\rm DFT}(t)$ for the residuals. The $s_{\epsilon,\rm DFT}(t)$ signal amplitude is far below the simulated correct value $A_2 = 2.0$. The blue dots representing the DFT model residuals are offset to the level of -3.0 and show clear trends.

Our Model 4 simulated time series analysis succeeds for the DCM and fails for the DFT.

3.5 Model 5

The mathematical time series equation

$$\begin{aligned} g(t) &= (A_1/2)\cos\left[\frac{2\pi(t - t_{1,\max,1})}{P_1}\right] \\ &+ (A_2/2)\cos\left[\frac{2\pi(t - t_{2,\max,1})}{P_2}\right] \\ &+ M_0 + M_1 T + M_2 T^2 \end{aligned} \quad (43)$$

for our fifth model is the same as for Model 4 (Equation 42). However, this new Model 5 differs from the earlier Model 4 because we use totally different P_1, P_2, A_1,A_2, $t_{1,\max,1}$, $t_{2,\max,1}$, M_0, M_1 and M_2 values (Table 5, Column 1). The simulated time series is "too short" for both P_1 and P_2 periods (Equation 35). There is "a trend" $p(t)$ (Equation 36). The signal frequencies f_1 and f_2 are "too close" (Equation 37). The three main reasons that can cause the failure of DFT analysis are present. We perform the DCM and DFT time series analysis between $P_{\rm min} = P_1/3 = 0.47$ and $P_{\rm max} = 3P_1 = 4.20$.

Model 5 has DCM orders $K_1 = 2$, $K_2 = 1$ and $K_3 = 2$. We give the DCM analysis results for different n and SN combinations in Table 5 (Columns 2-4). The DCM fails to detect the correct P_1, P_2, ..., M_1 and M_2 values for the the lowest $n = 10\,000$ and SN = 1 000 combination (Table 5, Column 2). This shows that the DCM can fail, just like any other time series analysis method, if the quality of data is too low. Due to the WD-effect, the DCM results for higher n and SN combinations are correct (Table 5, Columns 3-4).

We show the DCM analysis results for one sample of Model 5 simulated time series (Figures 5a-d: $n = 10\,000$ and SN = 10 000). The long and short DCM searches give $P_1 = 1.407$ and $P_2 = 1.917$ (Figure 5a: diamonds), and $P_1 = 1.393$ and $P_2 = 2.024$ (Figure 5b: Diamonds). The continuous line denoting the model $g(t)$ is totally covered by the black dots of y_i data (Figure 5c). Therefore, we use white colour to highlight this DCM model $g(t)$ line. The detected polynomial trend $p(t)$ coefficients $M_0 = 0.81 \pm 0.16$, $M_1 = 0.230 \pm 0.013$ and

Table 5 Model 5. DCM analysis between $P_{\min} = 0.47$ and $P_{\max} = 4.20$. Notations as in Table 1.

(1) Model 5	(2) $n = 10\ 000$ SN = 1 000	(3) $n = 10\ 000$ SN = 10 000	(4) $n = 100\ 000$ SN = 10 000
$P_1 = 1.4$	1.404 ± 0.046	1.393 ± 0.014	1.4019 ± 0.0035
$A_1 = 2.0$	2.27 ± 0.45	1.97 ± 0.13	1.994 ± 0.029
$t_{1,\min,1} = 1.1$	1.129 ± 0.023	1.1030 ± 0.0028	1.09766 ± 0.00082
$t_{1,\max,1} = 0.4$	0.4267 ± 0.0012	0.4066 ± 0.0059	0.3967 ± 0.0020
$P_2 = 1.9$	5.24 ± 0.83	2.023 ± 0.094	1.854 ± 0.023
$A_2 = 2.0$	5 ± 24	2.35 ± 0.34	1.918 ± 0.074
$t_{2,\min,1} = 1.55$	3.23 ± 0.50	1.599 ± 0.030	1.5292 ± 0.0066
$t_{2,\max,1} = 0.6$	0.61 ± 0.34	0.587 ± 0.030	0.6024 ± 0.0064
$M_0 = 1.0$	-2 ± 12	0.81 ± 0.16	1.055 ± 0.035
$M_1 = 0.25$	-1.27 ± 0.53	0.230 ± 0.013	0.2590 ± 0.0029
$M_2 = 0.5$	3.8 ± 1.8	0.596 ± 0.084	0.470 ± 0.020
Data file	`Model5n10000SN1000.dat`	`Model5n10000SN10000.dat`	`Model5n100000SN10000.dat`
Control file	`dcmModel5n10000SN1000.dat`	`dcmModel5n10000SN10000.dat`	`dcmModel5n100000SN10000.dat`

Table 6 Model 6. DCM analysis between $P_{\min} = 0.053$ and $P_{\max} = 0.480$. Notations as in Table 1.

(1) Model 6	(2) $n = 50$ SN = 10	(3) $n = 50$ SN = 50	(4) $n = 100$ SN = 100
$P_1 = 0.16$	0.15966 ± 0.00020	0.160022 ± 0.000021	0.160024 ± 0.000019
$A_1 = 2.0$	$1.983 \pm \pm 0.041$	2.0031 ± 0.0086	2.0013 ± 0.0032
$t_{1,\min,1} = 0.0979$	0.09868 ± 0.00057	0.097933 ± 0.000096	0.097934 ± 0.000079
$t_{1,\min,2} = 0.0225$	0.02283 ± 0.00070	0.022563 ± 0.000077	0.022403 ± 0.000071
$t_{1,\max,1} = 0.1421$	0.14241 ± 0.00064	0.142099 ± 0.000079	0.142100 ± 0.000077
$t_{1,\max,2} = 0.0575$	0.05827 ± 0.00057	0.05761 ± 0.00011	0.057448 ± 0.000074
$M_0 = 0$	0.003 ± 0.012	-0.0055 ± 0.0026	0.00176 ± 0.00088
Data file	`Model6n50SN10.dat`	`Model6n50SN50.dat`	`Model6n100SN100.dat`
Control file	`dcmModel6n50SN10.dat`	`dcmModel6n50SN50.dat`	`dcmModel6n100SN100.dat`

$M_2 = 0.596 \pm 0.084$ are correct (Figure 5c: dashed black line). We show the detrended DCM model $g(t) - p(t)$ (white continuous line), the detrended data $y(t_i) - p(t_i)$ (black dots), the signal $h_1(t)$ (red thick continuous line) and the signal $h_2(t)$ (blue thin continuous line) in Figure 5d. The DCM residuals (blue dots) are offset to the level of -1.5. These blue dots appear black because there are 10 000 of them. For obvious reasons, the -1.5 level of these residuals is highlighted by a white dotted line. The DCM model residuals are stable and show no trends.

The DFT detects the wrong periods for the original data (Figure 5e: diamond at $P_1 = 0.592$) and for the residuals (Figure 5f: diamond at $P_2 = 1.012$). The DFT estimates for the $p(t)$ trend, $M_0 = 2.758$, $M_1 = 0.091$ and $M_2 = -2.120$, are also wrong (Figure 5g: dashed black line). The DFT model $g_{\mathrm{DFT}}(t)$ black continuous line deviates from the black dots of data y_i, especially in the end of the simulated data sample. The black dots denoting the detrended data $y(t_i) - p_{\mathrm{DFT}}(t_i)$ and the continuous black line denoting the detrended DFT model $g_{\mathrm{DFT}}(t) - p_{\mathrm{DFT}}(t)$ are shown in Figure 5h. The DFT model $g_{\mathrm{DFT}}(t)$ gives very low amplitudes for the pure sine signal $s_{\mathrm{y,DFT}}(t)$ (continuous thick red line) and pure sine signal $s_{\epsilon,\mathrm{DFT}}(t)$ (continuous thinner blue line). These amplitudes are far below the correct simulated values $A_1 = A_2 = 2$. The offset level for the DFT model residuals (blue dots) is -0.5. These residuals show strong trends.

The DCM analysis succeeds for Model 5 simulated time series, but the DFT analysis fails.

3.6 Model 6

Our sixth time series simulation model is

$$g(t) = c_1 \cos\left[\frac{2\pi t}{P_1}\right] + c_2 \cos\left[\frac{4\pi(t - c_3)}{P_1}\right] + M_0, \tag{44}$$

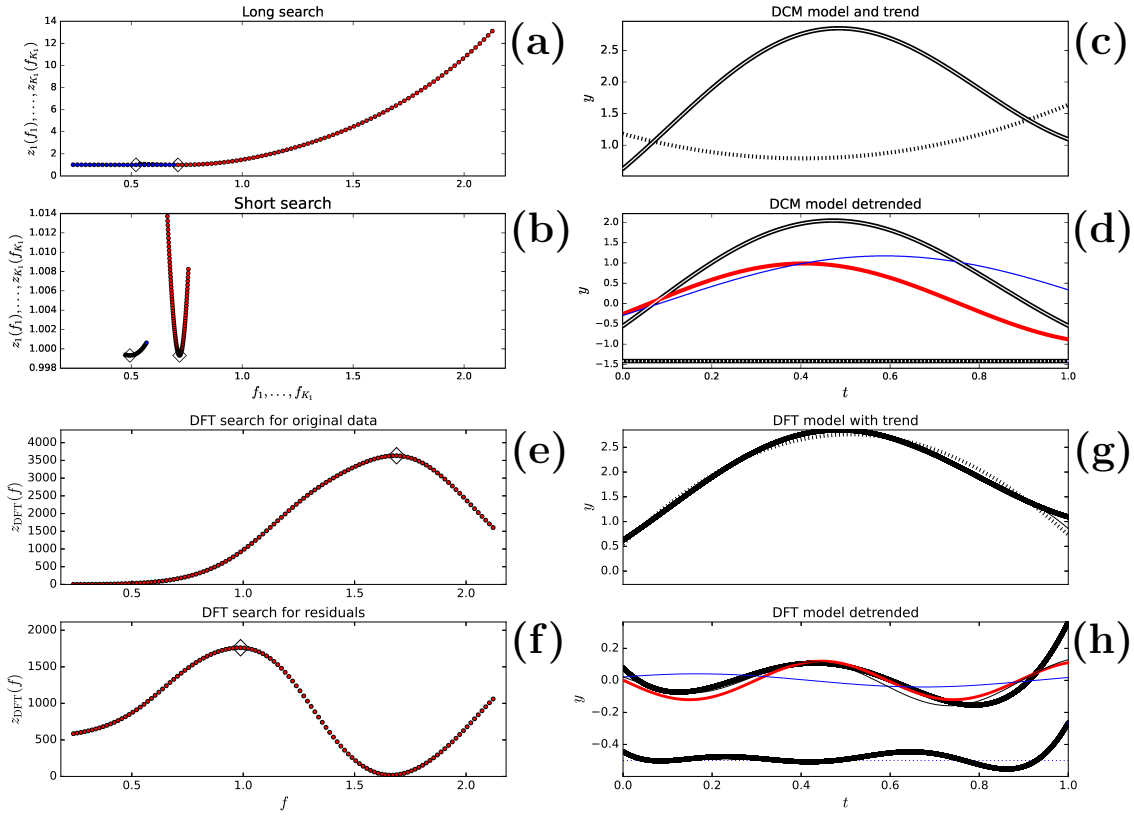

Fig. 5 Model 5 (Table 5: $n = 10\ 000$, SN $= 10\ 000$ simulation). Notations as in Figure 3. Best periods (diamonds) are explained in Section 3.5.

Table 7 Model 7. DCM analysis between $P_{\min} = 0.4$ and $P_{\max} = 3.6$. Notations as in Table 1.

(1) Model 7	(2) $n = 100$ SN $= 5\ 000\ 000$	(3) $n = 1\ 000$ SN $= 1\ 000\ 000$	(4) $n = 10\ 000$ SN $= 1\ 000\ 000$
$P_1 = 1.2$	1.19917 ± 0.00070	1.20000 ± 0.00063	1.20000 ± 0.00024
$A_1 = 2.0$	2.20 ± 0.20	1.94 ± 0.14	1.990 ± 0.057
$t_{1,\min,1} = 0.6892$	0.6835 ± 0.0050	0.6900 ± 0.0043	0.6888 ± 0.0019
$t_{1,\min,2} = 0.1134$	0.1235 ± 0.0098	0.1160 ± 0.0067	0.1128 ± 0.0026
$t_{1,\max,1} = 1.0195$	1.0253 ± 0.0052	1.0176 ± 0.0039	1.0188 ± 0.0018
$t_{1,\max,2} = 0.3779$	0.3705 ± 0.0082	0.3804 ± 0.0050	0.3780 ± 0.0020
$P_2 = 1.4$	1.4050 ± 0.0048	1.3986 ± 0.00034	1.3997 ± 0.0014
$A_2 = 2.0$	1.871 ± 0.098	2.049 ± 0.095	2.009 ± 0.040
$t_{2,\min,1} = 0.9262$	0.938 ± 0.012	0.9231 ± 0.0090	0.9252 ± 0.0036
$t_{2,\min,2} = 0.2766$	0.260 ± 0.016	0.281 ± 0.012	0.2771 ± 0.0046
$t_{2,\max,1} = 1.3109$	1.3165 ± 0.0049	1.3105 ± 0.0039	1.3101 ± 0.0014
$t_{2,\max,2} = 0.5864$	0.5817 ± 0.0037	0.5860 ± 0.0042	0.5865 ± 0.0015
$M_0 = 1.0$	0.950 ± 0.050	1.011 ± 0.035	1.002 ± 0.013
$M_1 = 0.25$	0.2447 ± 0.0067	0.2548 ± 0.0039	0.2509 ± 0.0018
$M_2 = 0.5$	0.551 ± 0.050	0.488 ± 0.035	0.498 ± 0.014
Data file	`Model7n100SN5000000.dat`	`Model7n1000SN1000000.dat`	`Model7n10000SN1000000.dat`
Control file	`dcmModel7n100SN5000000.dat`	`dcmModel7n1000SN1000000.dat`	`dcmModel7n10000SN1000000.dat`

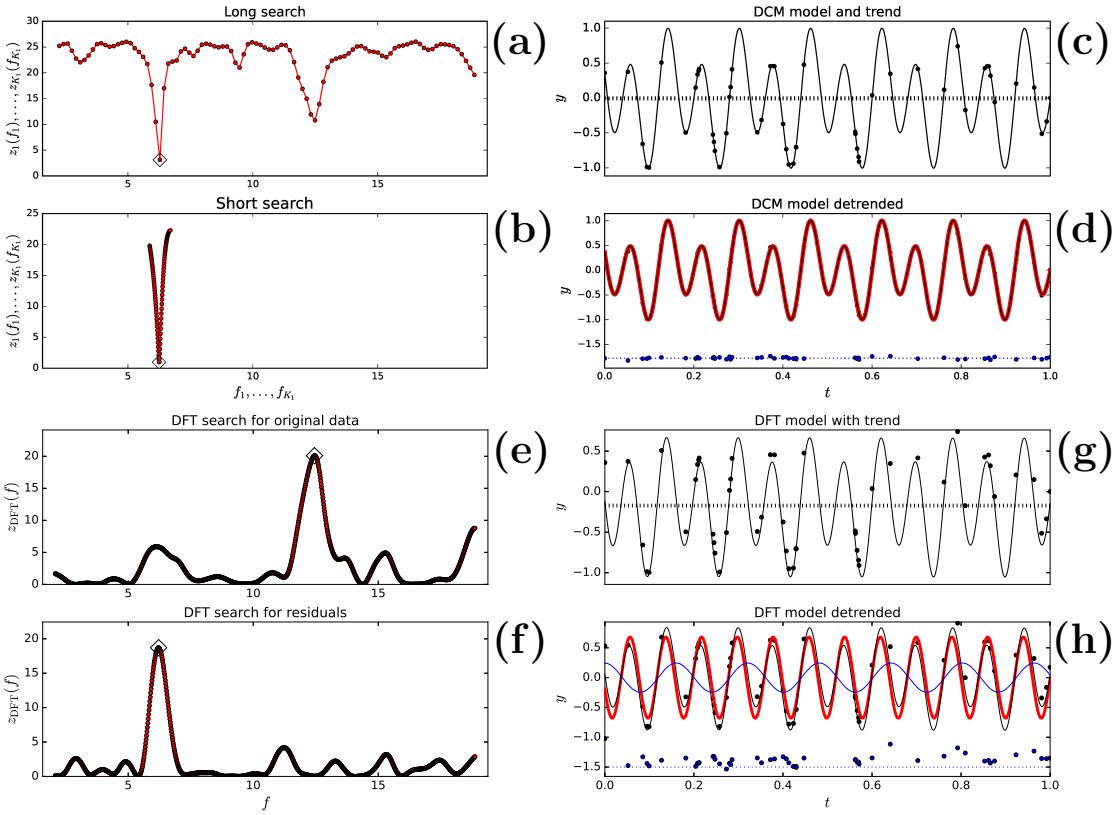

Fig. 6 Model 6 (Table 6: $n = 50$, SN = 50 simulation). Notations as in Figure 3. Best periods (diamonds) are explained in Section 3.6.

where $P = 0.16$ and $M_0 = 0$. The coefficients $c_1 = 0.3655$, $c_2 = 0.7310$ and $c_3 = 0.3000$ determine the A_1, $t_{1,\min,1}$, $t_{1,\min,2}$, $t_{1,\max,1}$ and $t_{1,\max,2}$ values given in Table 6 (Column 1). This simulated time series is not "too short" because $P_1 < \Delta T$ (Equation 35). There is no trend because $p(t) = M_0 = 0$ (Equation 36). This simulated time series contains only one signal (Equation 37). However, this simulation Model 6 is not a $K_2 = 1$ pure sine model (Equation 38). This double wave simulation model has two unequal minima and maxima. Its DCM orders are $K_1 = 1$, $K_2 = 2$ and $K_3 = 0$. We perform the DCM and DFT time series analysis between $P_{\min} = P_1/3 = 0.053$ and $P_{\max} = 3P_1 = 0.480$.

The DCM time series analysis results for different n and SN combinations are given in Table 6 (Columns 2-4). The DCM detects the correct P_1, A_1, $t_{1,\min,1}$ $t_{1,\min,2}$ $t_{1,\max,1}$ $t_{1,\max,2}$ and M_0 values even for the lowest $n = 50$ and SN = 10 combination. For increasing n and SN, the results for the model parameters converge to correct values (WD-effect).

We demonstrate the DCM analysis results for simulated time series having $n = 50$ and SN = 50 (Figures 6a-d). The long and short searches give $P_1 = 0.159$ (Figure 6a: diamond) and $P_1 = 0.160$ (Figure 6b: diamond). The DCM model $g(t)$ black line covers the black dots denoting the data y_i (Figure 6c). The "trend" at $p(t) = M_0 = -0.0055 \pm 0.0026$ is correct. The detrended model $g(t) - p(t)$ (black continuous line), the pure sine signal $h_1(t)$ (red thick continuous line) and the detrended data $y(t_i) - p(t_i)$ (black dots) are displayed in Figure 6d. The thick continuous red line stays under the thin continuous black line because $h_1(t) = g(t) - p(t)$. The DCM residuals (blue dots) are offset to the level of -1.8 (blue dotted line).

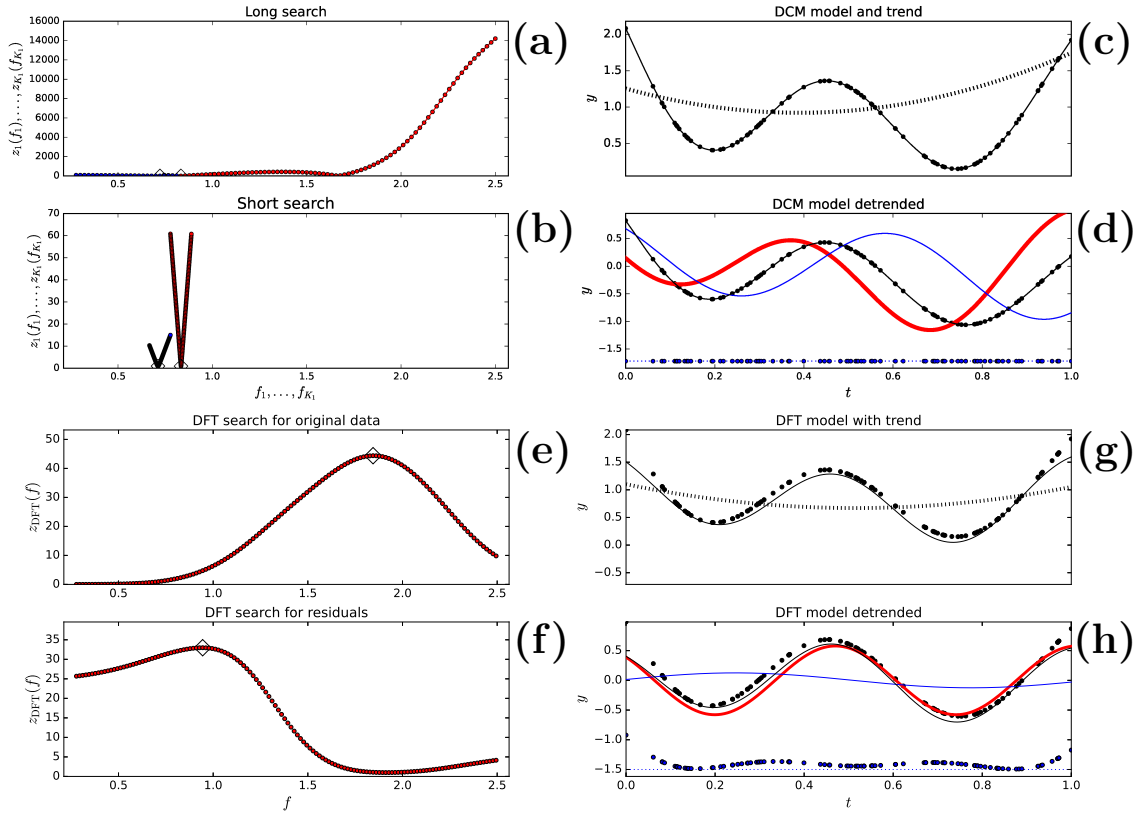

Fig. 7 Model 7 (Table 7: $n = 100$, SN = 5 000 000 simulation). Notations as in Figure 3. Best periods (diamonds) are explained in Section 3.7.

These residuals show no trends and their level is stable.

The DFT detects the wrong period $P_1 = 0.080$ (Figure 6e: Diamond). This results is exactly half of the correct simulated value $P_1 = 0.160$. The reason for this "detection" is that the double wave dominates because $c_2 = 2c_1$ in Model 6 (Equation 44). The DFT mean level estimate $M_0 = -0.172$ fails (Figure 6f: Dashed black line). The DFT analysis of the residuals $\epsilon_i = y(t_i) - [s_{\mathrm{y,DFT}}(t_i) + p_{\mathrm{DFT}}(t_i)]$ gives $P_2 = 0.161$, which is nearly equal to the correct simulated $P_1 = 0.160$ value. The black dots denoting the data y_i show minor deviations from continuous black line denoting the DFT model $g_{\mathrm{DFT}}(t)$ (Figure 6g). The detrended model $g_{\mathrm{DFT}}(t) - p_{\mathrm{DFT}}(t)$ (continuous black line), the detrended data $y(t_i) - p_{\mathrm{DFT}}(t_i)$ (black dots) and the signal $s_{\mathrm{y,DFT}}(t)$ (continuous thick red line) are shown in Figure 6h. The DFT analysis residuals (blue dots) are offset to the level of -1.5 (blue dotted line). These residuals display trends.

We conclude that the DFT "detects" the $P_1/2$ and P_1 periods, where P_1 is the correct simulated period value. However, the DFT two signal model is not the correct model for these Model 6 simulated time series, which contains only one signal. If the correct period is P and the correct model is a double wave ($K_2 = 2$), the DCM pure sine model ($K_1 = 1$) analysis can also give the values $P/2$ and P.

The DCM analysis succeeds for Model 6 simulated time series. The DFT analysis fails.

3.7 Model 7

In this section, we analyse our most complex time series. The seventh simulation model is

$$g(t) = c_1 \cos\left[\frac{2\pi t}{P_1}\right] + c_2 \cos\left[\frac{4\pi(t - c_3)}{P_1}\right] \quad (45)$$

$$+ \; c_4 \cos\left[\frac{2\pi t}{P_2}\right] + c_5 \cos\left[\frac{4\pi(t-c_6)}{P_2}\right]$$
$$+ \; M_0 + M_1 T + M_2 T^2$$

where $T = [2(t - t_{\rm mid})]/\Delta T$, $P_1 = 1.2$, $P_2 = 1.4$, $M_0 = 1$ $M_1 = 0.25$ and $M_2 = 0.5$. The coefficients $c_1 = 0.3687$, $c_2 = 0.7374$, $c_3 = 0.4000$, $c_4 = 0.3708$, $c_5 = 0.7416$ and $c_6 = 0.6000$ determine the A_1, $t_{1,\min,1}$, $t_{1,\min,2}$, $t_{1,\max,1}$, $t_{1,\max,2}$, A_2, $t_{2,\min,1}$, $t_{2,\min,2}$, $t_{2,\max,1}$ and $t_{2,\max,2}$ values given in Table 7 (Column 1). This simulated time series is "too short" because $\Delta T < P_1 < P_2$ (Equation 35). The parabolic $p(t)$ represents "a trend" (Equation 36). The signal frequencies f_1 and f_2 are "too close" because $\Delta f = f_1 - f_2 = 0.12 < f_0 = 1$ (Equation 37). The two $h_1(t)$ and $h_2(t)$ signals are not "pure sines" (Equation 38). These signals are double waves having two unequal minima and maxima. All four main reasons that can cause the failure of DFT analysis are present (Equations 35-38). Therefore, this simulated time series is the most complex one of all analysed seven time series. Our DCM and DFT time series analysis of Model 7 simulated time series is performed between $P_{\rm min} = P_1/3 = 0.4$ and $P_{\rm max} = 3P_1 = 3.6$.

The DCM orders of Model 7 are $K_1 = 2$, $K_2 = 2$ and $K_3 = 2$. This model has $\eta = 13$ free parameters (Equation 8). We give the DCM analysis results for different n and SN combinations in Table 7 (Columns 2-4). The DCM analysis results are displayed for one sample of Model 7 simulated time series (Figures 7a-d: $n = 100$ and SN = 5 000 000). Since there are $\eta = 13$ free model parameters, this $n = 100$ time series is quite small for time series analysis. The long and short DCM searches give $P_1 = 1.200$ and $P_2 = 1.385$ (Figure 7a: diamonds), and $P_1 = 1.199$ and $P_2 = 1.405$ (Figure 7b: Diamonds). The black $g(t)$ model line goes through all black y_i data dots (Figure 7c). The DCM detects the correct polynomial trend $p(t)$ coefficients $M_0 = 0.950 \pm 0.050$, $M_1 = 0.2447 \pm 0.0067$ and $M_2 = 0.551 \pm 0.050$ (Figure 7c: dashed black line). The detrended DCM model $g(t) - p(t)$ (white continuous line), the detrended data $y(t_i) - p(t_i)$ (black dots), the signal $h_1(t)$ (red thick continuous line) and the signal $h_2(t)$ (blue thin continuous line) are displayed in Figure 7d. The $n = 100$ residuals (blue dots) are offset to the level of -1.8 (dotted blue line). These DCM model residuals are small and their level is stable. These results confirm that the DCM time series analysis method can detect complex non-linear models ($\eta = 13$) from very small time series ($n = 100$), if the data are extremely accurate (SN = 5 000 000). These results demonstrate the power of WD-effect because both periods P_1 and P_2 are shorter than the sample window ΔT, the trend is a parabola and the signals are not pure sines.

The DFT time series analysis gives the wrong periods for the original data (Figure 7e: diamond at $P_1 = 0.542$ and for the residuals (Figure 7f: diamond at $P_2 = 1.059$. The DFT also gives wrong $p(t)$ trend coefficients $M_0 = 0.670$, $M_1 = -0.025$ and $M_2 = 0.406$ (Figure 7g: dashed black line). The DFT model $g_{\rm DFT}(t)$ (black continuous line) deviates from the data y_i (black dots). This deviation is largest at the beginning and the end of the simulated time series. In our Figure 7h, the black dots denote the detrended data $y(t_i) - p_{\rm DFT}(t_i)$, the continuous black line denotes the detrended DFT model $g_{\rm DFT}(t) - p_{\rm DFT}(t)$ and the continuous thick red line denotes pure sine signal $s_{\rm y,DFT}(t)$ detected from the original data. The DFT model $g_{\rm DFT}(t)$ gives very low amplitude for the pure sine signal $s_{\epsilon,\rm DFT}(t)$ detected from the residuals (continuous thinner blue line). The correct simulated peak to peak amplitude for this second signal is much larger, $A_2 = 2$. The DFT model residuals (blue dots) are offset to -1.5 level and show strong trends.

The DCM analysis succeeds for Model 7 simulated time series. The DFT analysis fails, as predicted by Equations 35-38.

3.8 Summary of simulations

For all seven simulated time series, all DCM analyses succeed and all DFT analyses fail (Sections 3.1 - 3.7).

The DCM detects exactly the correct signal(-s) and trend. Clearly, the AL6-AL10 do not constrain our DCM. We do admit that the simulated n and SN values of Models 2, 5 and 7 are extreme and unrealistic for most cases of real data. However, those model simulations are necessary for demonstrating the WD-effect. We could keep on adding complexity by simulating a larger number of signals (K_1), more complex signal shapes (K_2) and/or higher order polynomial trends (K_3). The WD-effect ensures that the DCM would detect those complex models when the sample size (n)

and/or the signal to noise (SN) increase. The other way round, the correct model can be found *if* the correct DCM model combination K_1, K_2 and K_3 is tested. Furthermore, totally wrong K_1, K_2 and K_3 combinations give unstable DCM models ("UM").

The analyses of seven simulated time series expose the DFT weaknesses. Every analysis fails! The AL6-AL10 are not the only causes for these failures. In general, the solutions for one-dimensional time series analysis models are ill-posed, if the data contain an unknown trend and/or more than one signal. For example, the detrending of simulated data misleads DFT even when the correct K_3 value is known (Sections 3.1-3.7). This correct K_3 value can be unknown for real data. The combination of detrending and iterative pre-whitening can fail at any stage. Even if the detrending succeeds, the leakage can cause the detection of wrong first signal frequency. This corrupts the model residuals and the whole analysis. For any one-dimensional time series analysis method, like the DFT, the search for many signals superimposed on an unknown trend can fail when the combination of detrending and pre-whitening is applied.

Our DCM outperforms the esteemed DTF. The DCM detects the correct signal(s) and trend, but DFT does not. For the sunspot data, the DCM detects signals superimposed on a constant trend (Jetsu 2025). The simulations presented here support the conclusion that those solar signals and the trend are correct. This would explain why the DCM is the first time series analysis method that detects Jupiter's exact $11.^{\mathrm{y}}86$ period in the sunspot record. The one-dimensional time series analysis methods, like DFF, have failed to find exactly correct periods.

3.9 Best model

The K_1, K_2 and K_3 orders of the best model are not necessarily known when some time series analysis method is applied to the real time series. We know a priori the best DCM and DFT model orders for the simulated time series of Models 1-7 (Sections 3.1-3.7). It could be argued that our DCM analysis succeeds only for this reason.

All alternative K_1, K_2 and K_3 order models are nested. For example, the *simple* one signal $K_1 = 1$ model $g_1(t)$ is a special case of the *complex* two signal $K_1 = 2$ model $g_2(t)$ having $A_2 = 0$. We use the Fisher-test to compare any pair of simple $g_1(t)$ and complex $g_2(t)$ models. The model parameters (Equation 14: R_1, R_2), (Equation 15: χ_1,χ_2) and (Equation 8: $\eta_1 < \eta_2$) give the test statistic of Fisher-test

$$F_R = \left(\frac{R_1}{R_2} - 1\right)\left(\frac{n - \eta_2 - 1}{\eta_2 - \eta_1}\right) \quad (46)$$

$$F_\chi = \left(\frac{\chi_1^2}{\chi_2^2} - 1\right)\left(\frac{n - \eta_2 - 1}{\eta_2 - \eta_1}\right). \quad (47)$$

The Fisher-test is based on the null hypothesis

> H_0: *"The complex model $g_2(t)$ does not provide a significantly better fit to the data than the simple model $g_1(t)$."*

Under this hypothesis, the F_R and F_χ test statistic parameters have an F distribution with $\nu_1 = \eta_2 - \eta_1$ and $\nu_2 = n - \eta_2$ degrees of freedom (Draper and Smith 1998). The critical level $Q_F = P(F_R \geq F)$ or $Q_F = P(F_\chi \geq F)$ is the probability that F_R or F_χ exceeds the numerical value F. If

$$Q_F < \gamma_F = 0.001, \quad (48)$$

we reject the H_0 hypothesis, which means that the complex $g_2(t)$ model is better than the simple $g_1(t)$ model. The pre-assigned significance level $\gamma_F = 0.001$ represents the probability that we falsely reject the H_0 hypothesis when it is in fact true.

Larger F_R or F_χ values have smaller Q_F critical levels. Hence, the probability for the H_0 hypothesis rejection increases when F_R or F_χ increases. If the number of complex model free parameters η_2 increases, the R_2 or χ_2^2 values decrease. This increases the F_R or F_χ values because the terms $(R_1/R_2 - 1)$ or $(\chi_1^2/\chi_2^2 - 1)$ increase (Equations 46 and 47: first terms). At the same time, the $(n - \eta_2 - 1)/(\eta_2 - \eta_1)$ penalty term decreases (Equations 46 and 47: second terms), and this decreases the F_R or F_χ values. This penalty term prevents the use of too high η_2 values (too complex models).

Here, we illustrate how the Fisher-test finds the best model from a group of numerous alternative nested DCM models. The Fisher-test is used to find the best model for the simulated time series of Model 1 combination $n = 100$ and

Table 8 Fisher tests. DCM analysis of Model 1 simulated time series ($n = 100, \mathrm{SN} = 100$, data file `Model1n100SN100`). (1) Simple model: Model number $\mathcal{M}$, model g_{K_1,K_2,K_3}, number of free parameters η, chi-square χ^2 and control file name. (2-8) Complex model: Hypothesis H_0 rejected $\uparrow$ (Complex model better), Hypothesis H_0 not rejected $\leftarrow$ (Simple model better), test statistic F and critical level Q_F.

(1)	(2)	(3)	(4)	(5) Complex model	(6)	(7)	(8)
Simple model	$\mathcal{M}$=2	$\mathcal{M}$=3	$\mathcal{M}$=4	$\mathcal{M}$=5	$\mathcal{M}$=6	$\mathcal{M}$=7	$\mathcal{M}$=8
$\mathcal{M}$=1	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$
$g_{1,1,-1}$	$F = 733$	$F = 367$	$F = 242$	$F = 242$	$F = 180$	$F = 142$	$F = 121$
$\eta = 3, \chi^2 = 738$	$Q_\mathrm{F} < 10^{-16}$	$Q_\mathrm{F} < 10^{-16}$	$Q_\mathrm{F} < 10^{-16}$	$Q_\mathrm{F} < 10^{-16}$	$Q_\mathrm{F} < 10^{-16}$	$Q_\mathrm{F} < 10^{-16}$	$Q_\mathrm{F} < 10^{-16}$
`dcmModel1K11-1.dat`							
$\mathcal{M}$=2	-	$\leftarrow$	$\leftarrow$	$\leftarrow$	$\leftarrow$	$\leftarrow$	$\leftarrow$
$g_{1,1,0}$	-	$F = 1.0451$	$F = 0.5404$	$F = 0.5159$	$F = 0.3553$	$F = 0.2784$	$F = 0.7686$
$\eta = 4, \chi^2 = 84.7422$	-	$Q_\mathrm{F} = 0.309$	$Q_\mathrm{F} = 0.584$	$Q_\mathrm{F} = 0.599$	$Q_\mathrm{F} = 0.785$	$Q_\mathrm{F} = 0.891$	$Q_\mathrm{F} = 0.575$
`dcmModel1K110.dat`							
$\mathcal{M}$=3	-	-	$\leftarrow$	$\leftarrow$	$\leftarrow$	$\leftarrow$	$\leftarrow$
$g_{1,1,1}$	-	-	$F = 0.0464$	$F = -0.0021$	$F = 0.0213$	$F = 0.0336$	$F = 0.7027$
$\eta = 5, \chi^2 = 83.8104$	-	-	$Q_\mathrm{F} = 0.830$	$Q_\mathrm{F} = 1$	$Q_\mathrm{F} = 0.979$	$Q_\mathrm{F} = 0.992$	$Q_\mathrm{F} = 0.592$
`dcmModel1K111.dat`							
$\mathcal{M}$=4	-	-	-	$\leftarrow$	$\leftarrow$	$\leftarrow$	$\leftarrow$
$g_{1,1,2}$	-	-	-	No test[1]	$F = -0.0033$	$F = 0.0277$	$F = 0.9215$
$\eta = 6, \chi^2 = 83.7686$	-	-	-		$Q_\mathrm{F} = 1$	$Q_\mathrm{F} = 0.973$	$Q_\mathrm{F} = 0.434$
`dcmModel1K112.dat`							
$\mathcal{M}$=5 UM : AD	-	-	-	-	$\leftarrow$	$\leftarrow$	$\leftarrow$
$g_{2,1,-1}$	-	-	-	-	$F = 0.0447$	$F = 0.0515$	$F = 0.9377$
$\eta = 6, \chi^2 = 83.8123$	-	-	-	-	$Q_\mathrm{F} = 0.833$	$Q_\mathrm{F} = 0.950$	$Q_\mathrm{F} = 0.426$
`dcmModel1K21-1.dat`							
$\mathcal{M}$=6 UM : IF, AD	-	-	-	-	-	$\leftarrow$	$\leftarrow$
$g_{2,1,0}$	-	-	-	-	-	$F = 0.0587$	$F = 1.3840$
$\eta = 7, \chi^2 = 83.7716$	-	-	-	-	-	$Q_\mathrm{F} = 0.809$	$Q_\mathrm{F} = 0.256$
`dcmModel1K210.dat`							
M=7 UM : IF, AD	-	-	-	-	-	-	$\leftarrow$
$g_{2,1,1}$	-	-	-	-	-	-	$F = 2.7081$
$\eta = 8, \chi^2 = 83.7176$	-	-	-	-	-	-	$Q_\mathrm{F} = 0.103$
`dcmModel1K211.dat`						-	
$\mathcal{M}$=8 UM : IF, AD	-	-	-	-	-	-	-
$g_{2,1,2}$	-	-	-	-	-	-	-
$\eta = 9, \chi^2 = 81.2721$	-	-	-	-	-	-	-
`dcmModel1K212.dat`						-	

SN = 100 (Table 1: column 4). In other words, we assume that the correct DCM model orders K_1, K_2 and K_3 are unknown, which can be the case for real time series. The eight tested models contain one or two signals, and no trend or a constant trend or a linear trend or a parabolic trend. We compare all these eight models $\mathcal{M}$=1-8 against each other (Table 8). The special model number notation"$\mathcal{M}$" is used because the notations "M_0, ..., M_{K_3}" have already been reserved for the $p(t)$ trend.

Model $\mathcal{M}$=2 has the known correct Model 1 orders $K_1 = 1$, $K_2 = 1$ and $K_3 = 0$. For example, the Fisher-test between the simple model $\mathcal{M}$=1 ($\eta = 3$, $\chi^2 = 738$) and the complex model $\mathcal{M}$=2 ($\eta = 4$, $\chi^2 = 84.7422$) gives $F = 733$ (Equation 47). The critical level[4] for this very large F value is extremely significant, $Q_F < 10^{-16}$. This means that the H_0 hypothesis must be rejected, and the complex $\mathcal{M}$=2 model is certainly better than the simple $\mathcal{M}$=1 model (Equation 48). The upward arrow "↑" in Table 8 indicates that $\mathcal{M}$=2 model is a better model than $\mathcal{M}$=1 model. A closer look at Table 8 reveals that $\mathcal{M}$=2 model is better than all other models because the "↑" and "←" arrows of all other models point toward $\mathcal{M}$=2 model. There is no need to test models having more than two signals because all two signal $\mathcal{M}$=5-8 models are already unstable (Table 8:"UM"). The Fisher-test finds the correct DCM model for Model 1 simulated time series.

In the above example, the Fisher-test finds the correct number of signals (K_1) and the correct trend (K_3) for the pure sine signal alternative ($K_2 = 1$). We do not test the double wave signal alternative ($K_2 = 2$) against the pure sine signal alternative ($K_1 = 1$) because the number of tested models would increase from 8 to 16, and Table 8 would become excessively large. One example of testing the $K_2 = 2$ signal models against each other can be found in Jetsu (2025, Table S7).

We conclude that the best model for the real time series can be found by applying the Fisher-test to any arbitrary number of different nested DCM or DFT models.

[4]The highest achievable accuracy for the computational f.cdf subroutine in scipy.optimize python library is 10^{-16}.

3.10 Significance estimates

Jetsu (2020) or Jetsu (2025) gave no signal significance estimate for the first detected period. Here, we use the Fisher-test for this purpose. The one signal model is the complex model. The logical simple model alternative is $g(t) = m \equiv$ white noise having standard deviation s. However, white noise is not the only possible alternative simple model. The other possible nested no signal DCM polynomial models are

$$g(t) = p(t, K_3) \tag{49}$$

where $h(t) = 0$, $K_3 = 0, 1, 2...$ The $K_3 = 0$ polynomial $g(t) = M_0$ represents white noise. The Fisher-test gives the critical level Q_F for rejecting the H_0 hypothesis when this hypothesis is in fact true (Equation 48). Therefore, this Q_F value represents the probability of false signal detection. In cases $Q_F < 10^{-16}$, the signal detection is absolutely certain.

We use the Model 1 simulated time series combination $n = 100$ and SN = 100 to demonstrate the significance estimation for the first detected period. The correct model for the first detected signal is $g_{1,1,0}(t)$. The Fisher-test is used to compare this correct $g_{1,1,0}(t)$ model to different polynomial models $g(t) = p(t, K_3)$ having $K_3 = 0, 1, .., 7$ (Table 9: Equation 49). The $\mathcal{M}$=1, 2 and 3 models have less free parameters than the correct $g_{1,1,0}(t)$ model, but the χ^2 values of these three polynomial models are so large that the Fisher-test is merely a formality. The $\mathcal{M}$=4 model has the same number of free parameters as the correct $g_{1,1,0}(t)$ model, but the comparison of χ^2 values reveals that this third order $p(t)$ polynomial is not the correct model for the time series. The next model $\mathcal{M}$=5 has more free parameters than the correct $g_{1,1,0}(t)$ model. This fourth order $p(t)$ polynomial $\mathcal{M}$=5 model must be rejected because it has a larger χ^2 value than the correct $g_{1,1,0}(t)$ model. The critical levels $Q_F \gg \gamma = 0.001$ for the remaining $\mathcal{M}$=6, 7 and 8 polynomial models are so large that these models must also be rejected. The results in Table 9 confirm that the one signal model is better than any polynomial model in Equation 49. Hence, the analysed Model 1 simulated time series must contain at least the $h_1(t)$ pure sine signal. For the constant, linear or

parabolic $p(t)$ trend alternatives, the significance for this $h_1(t)$ signal is $Q_{\rm F} < 10^{-16}$.

After the detection of the first strongest signal, the Fisher-test critical level $Q_{\rm F}$ values increase for the next detected weaker signals. In other words, the signal significances decrease. Typical examples can be found in Jetsu 2025 (Tables S5-S16). No new signals are detected when $Q_{\rm F} > \gamma = 0.001$ because the critical level exceeds the preassigned significance level and the H_0 hypothesis is no longer rejected. For the Model 1 simulated $n = 100$ and SN = 100 combination, Table 8 shows that this time series contains only one signal.

For this Model 1 time series, there is no need to discuss the DFT significance estimates (Horne and Baliunas 1986, their Equation 22) because this method fails to detect the correct period.

We conclude that the Fisher-test identifies the correct DCM model (Section 3.9), as well as gives the signal significance $Q_{\rm F}$ estimates (Section 3.10).

Table 9 Fisher tests for Model 1 $n = 100$, SN = 100 simulated time series (electronic data file `Model1n100SN100`). (1) Polynomial model (Equation 49). (2) DCM model $g_{1,1,0}$. Note that $\mathcal{M}$=1-3 polynomials represent simple models, and $\mathcal{M}$=5-8 polynomials represent complex models. Otherwise as in Table 8.

(1)	(2)
	$g_{1,1,0}$
Polynomial	$\eta = 4, \chi^2 = 84.7422$
$\mathcal{M}$=1	↑
$g(t) = p(t, K_3 = 0)$	$F = 63218$
$\eta = 1$, $\chi^2 = 169260$	$Q_{\rm F} < 10^{-16}$
$\mathcal{M}$=2	↑
$g(t) = p(t, K_3 = 1)$	$F = 54366$
$\eta = 2$, $\chi^2 = 97076$	$Q_{\rm F} < 10^{-16}$
$\mathcal{M}$=3	↑
$g(t) = p(t, K_3 = 2)$	$F = 1945$
$\eta = 3$, $\chi^2 = 1820$	$Q_{\rm F} < 10^{-16}$
$\mathcal{M}$=4	↑
$g(t) = p(t, K_3 = 3)$	No test[1]
$\eta = 4$, $\chi^2 = 469$	
$\mathcal{M}$=5	↑
$g(t) = p(t, K_3 = 4)$	$F = -0.3718$
$\eta = 5$, $\chi^2 = 85.0787$	$Q_{\rm F} = 1$
$\mathcal{M}$=6	↑
$g(t) = p(t, K_3 = 5)$	$F = 0.3403$
$\eta = 6$, $\chi^2 = 84.1265$	$Q_{\rm F} = 0.712$
$\mathcal{M}$=7	↑
$g(t) = p(t, K_3 = 6)$	$F = 0.3524$
$\eta = 7$, $\chi^2 = 83.7794$	$Q_{\rm F} = 0.787$
$\mathcal{M}$=8	↑
$g(t) = p(t, K_3 = 7)$	$F = 0.9394$
$\eta = 8$, $\chi^2 = 81.3819$	$Q_{\rm F} = 0.445$

Note: No Fisher-test for $\mathcal{M}$=4 ($\eta_1 = \eta_2 = 4$).

3.11 Ill-posed problem

The solution for the non-linear DCM model is an ill-posed problem (Equations 1-5). We present a computational statistical solution that fulfils the C_1, C_2 and C_3 conditions of a well-posed problem. The DCM model is just one special case of a non-linear model. Our technique can be applied to solve other non-linear models: Use the free parameters that make the model non-linear (Equation 9: $\boldsymbol{\beta}_I$) for solving the remaining other free parameters (Equation 10: $\boldsymbol{\beta}_{II}$).

3.11.1 Existence (C_1)

Fourier (1822) transformed the original function into the frequency domain. The modern DFT time series analysis method transforms the original time series into the frequency domain and gives the best frequency for a pure sine model. Gauß (1821) presented the LS method, which minimises the differences between the data and the linear model. Our DCM does the same by testing a large number of linear models. The data spacing, even or uneven, is irrelevant for these models. For every chosen DCM model, the total number of tested linear models is

$$n_{\rm Lin} = \binom{n_{\rm L}}{K_1} + (1 + n_{\rm B}) \times \binom{n_{\rm S}}{K_1}, \qquad (50)$$

where the number of tested long and short search frequencies is $n_{\rm L}$ and $n_{\rm S}$, respectively. The number of bootstrap samples is $n_{\rm B}$. In other words, we solve this ill-posed problem by using brute computational force. If the time series contains only zero mean white noise, the Gauß-Markov theorem ensures that a LS fit solution *exists* for every tested frequency combination.

3.11.2 Uniqueness (C_2)

We make the most of the Gauß-Markov theorem. When the numerical values of the tested frequencies $\boldsymbol{\beta}_I$ (Equation 9) are fixed, the model becomes linear and the solution for the other free parameters $\boldsymbol{\beta}_{II}$ is *unique* (Equation 10). All possible $\boldsymbol{\beta}_I$ frequency combinations are tested (Equation 12). For every tested frequency combination $\boldsymbol{\beta}_I$,

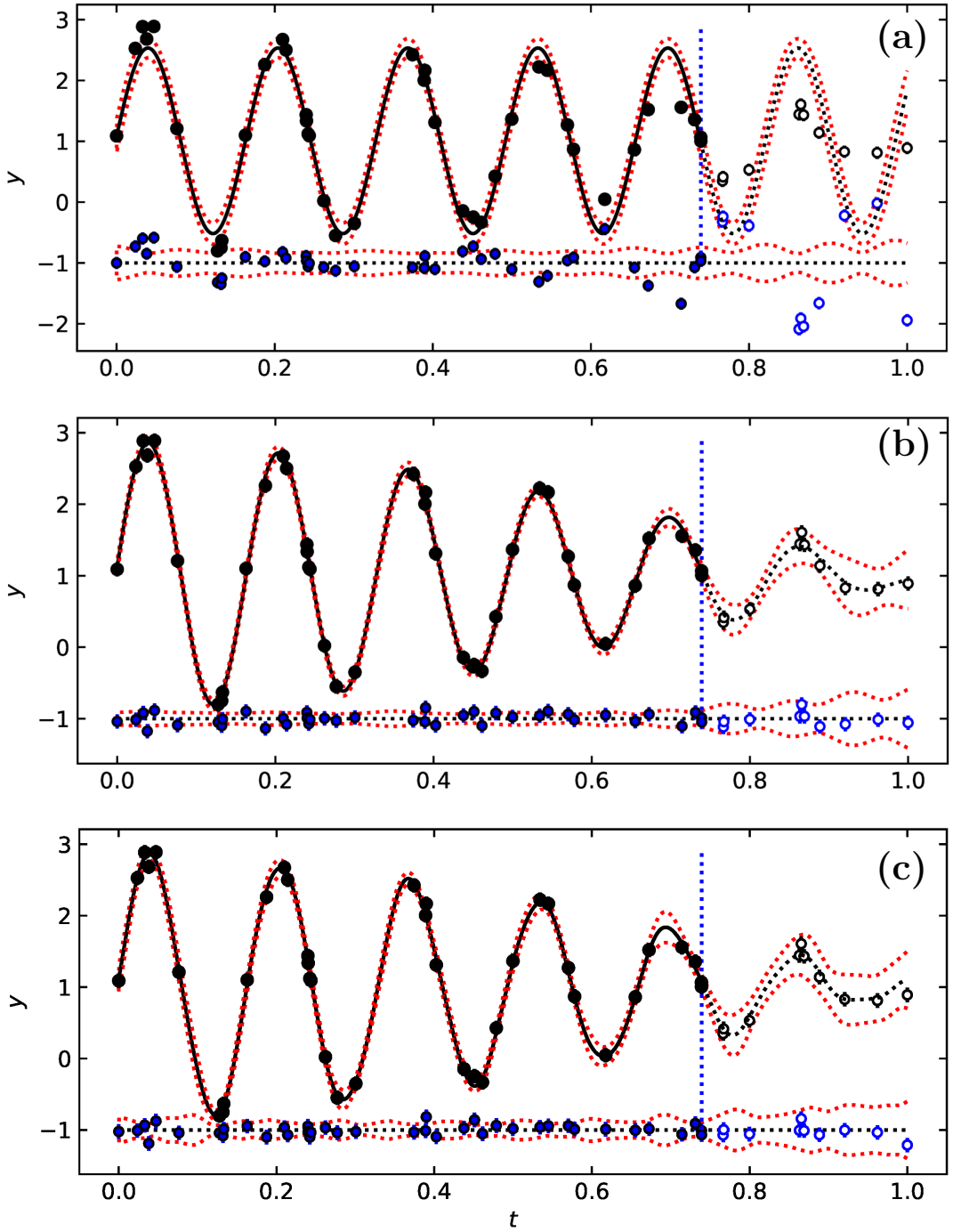

Fig. 8 DCM model forecast (Model 3 simulated time series for combination $n = 50$ and SN = 10: first forty observations are forecasting data). (a) One signal model $g_{1,1,0}$ results. Forecasting data are t_i, y_i and σ_i ($n = 40$ black dots) and forecasted data are ten last observations t'_i, y'_i and σ'_i ($n = 10$ open black dots). Continuous black line of forecasting data model $g(t, \boldsymbol{\beta}_{\text{Fore}})$ ends to vertical dotted blue line, where dotted black line of forecasted model $g'(t, \boldsymbol{\beta}_{\text{Fore}})$ begins. Dotted red line denotes $\pm 3\sigma_{g(t)}$ errors of both models (Equation 30). Residuals of forecasting data model $\epsilon_i = y_i - g_i$ (blue dots), residuals of forecasted data model $\epsilon_i =' y'_i - g'_i$ (open blue dots) and $\pm 3\sigma$ errors of both models (red dotted line) are offset to level -1 (blue dotted line). (b) Two signal model $g_{2,1,0}$ results. Otherwise as in "a". (c) Three signal model $g_{3,1,0}$ results. Otherwise as in "a".

Table 10 DCM time series analysis between $P_{\min} = 0.53$ and $P_{\max} = 4.80$ for forecasting data (Model = 3 combination $n = 50$ and SN = 10 simulated data: first forty observations). (1) Model. (2) Forecasting data test statistic z (Equation 17). (3) Forecasted data test statistic z_{Fore} (Equation 53). (4) Data file. (5) Control file. (6) Figure where forecast is shown.

(1) Model	(2) z	(3) z_{Fore}	(4) Data file	(5) Control file	(6) Figure
$g_{1,1,0}$	2.27	8.61	`Model3n40SN10.dat`	`dcmModel3K110.dat`	8a
$g_{2,1,0}$	0.76	0.88	`Model3n40SN10.dat`	`dcmModel3K210.dat`	8b
$g_{3,1,0}$	0.70	0.91	`Model3n40SN10.dat`	`dcmModel3K310.dat`	8c

the linear model gives a *unique* value for the test statistic z (Equations 16 and 17). From all tested frequency combinations, we select the best frequency combination $\boldsymbol{\beta}_{\text{I,best}}$ which minimises z. The linear model for this best frequency combination $\boldsymbol{\beta}_{\text{I,best}}$ gives *unique* values for the remaining other free parameters $\boldsymbol{\beta}_{\text{II,best}}$. The only goal for the massive DCM search (Equation 50) is to find these *unique* initial free parameter values $\boldsymbol{\beta}_{\text{initial}} = [\boldsymbol{\beta}_{\text{I,best}}, \boldsymbol{\beta}_{\text{II,best}}]$ for the non-linear iteration that gives the *unique* final free parameter values $\boldsymbol{\beta}_{\text{final}}$ (Equation 22).

We use the Fisher-test to compare many different non-linear DCM models against each other. The selection criterion for the best model is *unique* (Equation 48). The best DCM model is not necessarily the correct model, if this correct model is not among the compared models. The correct model must be able to forecast the future and past data. We formulate the Forecast-test for alternative DCM models (Equation 53). The order of Fisher- and Forecast-tests can be reversed. However, the former uses all data, while the latter uses a subset of all data, the forecasting data. In ideal cases, both tests identify the same best and correct model, like in Section 3.12.

3.11.3 Stability (C_3)

The artificial bootstrap data sets (Equation 31) represent "small changes in the input data", while the bootstrap results for the model parameters represent "small chances in the solution" (Section 1: C_3 condition formulation). We routinely check the *stability* of these solutions (e.g. Jetsu 2025, Figure S5). The unstable models, where the model parameter changes are large, are rejected (Section 2.1: "UM" models).

There are additional signatures of *stability*. The z periodogram solution is unique for every tested $\boldsymbol{\beta}_I$ frequency combination. If these periodograms are continuous and their changes are not irregular (e.g., like in Figs. 1a-b), the DCM model solution is *stable* because it does not change by increasing the number of tested frequencies n_{L} and n_{S}. Furthermore, the solutions for all seven simulated time series converge when the n and SN values increase (Tables 1 - 7). The different n and SN combinations give the same *stable* DCM model solution.

We conclude that our computational statistical DCM model solution fulfils the C_1, C_2 and C_3 conditions of the solution for a well-posed problem (Sections 3.11.1 - 3.11.3).

3.12 Forecast

There are numerous techniques for forecasting a time series (e.g., Hamilton 1994; Hastie et al. 2001; Kazemi and Rodrigues 2025). The DCM model $g(t_i, \boldsymbol{\beta})$ can be used to forecast. We divide all data into the forecasting data and the forecasted data. The time points, the observations and the errors of these samples are

n forecasting data values t_i, y_i and σ_i
n' forecasted data values t'_i, y'_i and σ'_i

The DCM gives the best forecasting data model

$$g_i = g(t_i, \boldsymbol{\beta}_{\text{Fore}}), \tag{51}$$

where $\boldsymbol{\beta}_{\text{Fore}}$ are the free parameter values. The $t_{\text{mid,Fore}} = (t_n + t_1)/2$ and $\Delta T_{\text{Fore}} = t_n - t_1$ values are computed from the forecasting data time points t_i. The forecasting model values at any arbitrary time t can be obtained from the $\boldsymbol{\beta}_{\text{Fore}}$ and $t_{\text{mid,Fore}}$ ΔT_{Fore} values.

The forecasted data model values are

$$g'_i = g(t'_i, \boldsymbol{\beta}_{\text{Fore}}), \tag{52}$$

where $t_{\rm mid} = t_{\rm mid,Fore}$ and $\Delta T = \Delta T_{\rm Fore}$. We do not compute "new" $t_{\rm mid}$ and ΔT values from the forecasted data time points t'_i because the correct g'_i values are obtained only from the $\boldsymbol{\beta}_{\rm Fore}$, $t_{\rm mid,Fore}$ and $\Delta T_{\rm Fore}$ combination of the forecasting model (Equation 52). The n' forecasted data model residuals

$$\epsilon'_i = y'_i - g'_i$$

give the *forecasted data test statistic*

$$z_{\rm Fore} = \text{z test statistic for forecasted data } t'_i,\ y'_i \text{ and } \sigma'_i \text{ (Equation 16 or 17).} \quad (53)$$

This parameter $z_{\rm Fore}$ measures how well the forecast (Equation 52) obtained from the forecasting data works for the forecasted data. If the DCM detects a new signal from the forecasting data, there are two alternatives:

If this new signal is real, the $z_{\rm Fore}$ value of forecasted data decreases.
If this new signal is unreal, the $z_{\rm Fore}$ value of forecasted data increases.

This "Forecast-test" technique revealed at least five real signals in the sunspot record (Jetsu 2025).

We compute the $z_{\rm Fore}$ parameter value from the *known* forecasted data t'_i, y'_i and σ'_i values (Equation 53). Forecasts are possible even if all t'_i, y'_i or σ'_i forecasted data values are *unknown*. In this case, the t_i, y_i and σ_i values of all data can be used as forecasting data. The best DCM model $g(t_i, \boldsymbol{\beta}_{\rm Fore})$ for all data determines the correct $\boldsymbol{\beta}_{\rm Fore}$, $t_{\rm mid,Fore} = (t_n + t_1)/2$ and $\Delta T_{\rm Fore} = t_n - t_1$ combination of the forecasting model. The n' forecasted data time points t'_i can be created for any arbitrary chosen sample window $\Delta T' = t'_n - t'_i$. The n' forecasted $g'_i = g'_i(t'_i, \boldsymbol{\beta}_{\rm Fore})$ values are obtained from Equation 52. These g'_i values can be used, for example, to compute the *forecasted mean level*

$$m_{\rm Fore} = \frac{1}{n'} \sum_{i=1}^{n'} g'_i \quad (54)$$

during the chosen sample window $\Delta T' = t'_n - t'_1$. Jetsu (2025) used this $m_{\rm Fore}$ parameter to postdict the known $\Delta T'$ time intervals of prolonged solar activity minima, like the Maunder minimum between the years 1640 and 1720. Since the known mean level of all sunspot data was m, Jetsu (2025) used these three criteria for correct postdictions of past prolonged activity minima $\Delta T'$ time intervals:

If DCM detects a real new signal in all data, the $m_{\rm Fore}$ value decreases.
If DCM detects many real signals in all data, the $m_{\rm Fore}$ value falls below m.
If DCM detects an unreal new signal in all data, the $m_{\rm Fore}$ value increases.

We use Model 3 combination $n = 50$ and SN $= 10$ simulated data (Table 3, column 2) to illustrate the DCM forecasting technique. The black dots in Figures 8a-c are the first $n = 40$ forecasting data values t_i, y_i and σ_i. The open black dots denote the $n' = 10$ forecasted data values t'_i, y'_i and σ'_i. The continuous black line is the forecasting model $g(t)$ and the dotted black line is the forecasted model $g'(t)$. The red dotted line shows the $\pm 3\sigma$ errors of both models. The blue dots are the forecasting data $\epsilon_i = y_i - g_i$ residuals and the open blue dots are the forecasted data $\epsilon'_i = y'_i - g'_i$ residuals.

Model 3 is the sum of two $P_1 = 0.16$ and $P_2 = 0.17$ pure sine signals ($K_1 = 2, K_2 = 1$) superimposed on the constant mean level $M_0 = 1$ ($K_3 = 0$). We compute the z and $z_{\rm Fore}$ values for the $g_{1,1,0}$, $g_{2,1,0}$ and $g_{3,1,0}$ models (Table 10). These three models have the same $K_2 = 1$ and $K_3 = 0$ orders as Model 3, but their signal numbers $K_1 = 1, 2$ or 3 are different, the $g_{2,1,0}$ model being the correct simulation Model 3.

The correct $g_{2,1,0}$ model gives the smallest $z_{\rm Fore} = 0.88$ value (Table 10). Therefore, it is a better forecasting model than the $g_{1,1,0}$ and $g_{3,1,0}$ models. The one signal $g_{1,1,0}$ model forecast fails because the blue open circles denoting the forecasted data residuals $\epsilon'_i = y'_i - g'_i$ show large deviations from to the blue dotted line offset level of $\epsilon'_i = -1$ (Figure 8a). The two signal $g_{2,1,0}$ model forecast succeeds because all open blue dots denoting the forecasted data residuals stay close to the blue dotted line offset level $\epsilon'_i = -1$, as well as inside the red dotted $\pm 3\sigma$ model error limits (Figure 8b).

The three signal model $g_{3,1,0}$ periods are $P_1 = 0.0643 \pm 0.0018$, $P_2 = 0.1592 \pm 0.0012$ and $P_3 = 0.1716 \pm 0.0012$. The amplitudes of these pure sine

signals are $A_1 = 0.092 \pm 0.025$, $A_2 = 1.879 \pm 0.16$ and $A_3 = 1.82 \pm 0.15$. The periods and the amplitudes of the two strongest P_2 and P_3 signals are correct because they are the same as in Model 3 simulation (Table 3, column 1). Therefore, the $g_{3,1,0}$ forecast appears nearly as good as the $g_{2,1,0}$ forecast because the residuals ϵ'_i are close to the offset level of $\epsilon'_i = -1$ (blue open dots), and these residuals also stay inside the red dotted $\pm 3\sigma$ model error limits (Figure 8c). The amplitude $A_1 = 0.092$ of the third $P_1 = 0.0643$ signal is very low. Due this weak "unreal" P_1 signal, the $g_{3,1,0}$ model has a larger $z_{\rm Fore} = 0.91$ value than the correct $g_{2,1,0}$ model (Table 10). Finally, we note that the three signal model $g_{3,1,0}$ is not unstable ("UM") although the simulated time series contains only two signals.

For the forecasting data, the extreme Fisher-test critical levels $Q_{\rm F} < 10^{-13}$ confirm that the two signal $g_{2,1,0}$ and the three signal $g_{3,1,0}$ models are certainly better than the one signal $g_{1,1,0}$ model. The forecasting data parameters $n = 40$, $\eta_1 = 7$, $\eta_2 = 10$, $\chi_1 = 22.946$ and $\chi_2 = 19.790$ for the simple $g_{2,1,0}$ model and the complex $g_{3,1,0}$ model give the Fisher-test critical level $Q_{\rm F} = 0.28 > \gamma_{\rm F} = 0.001$. The $g_{2,1,0}$ model beats the $g_{3,1,0}$ model because the H_0 hypothesis is not rejected. Hence, the Fisher-test confirms that the $g_{2,1,0}$ model is the best model for the forecasting data. The best DCM model (Fisher-test, $g_{2,1,0}$) is also the correct DCM model (Forecast-test, $g_{2,1,0}$). The double-check works for this particular Model 3 simulated time series!

The DFT can not detect the correct frequencies for the forecasting data because the simulated frequencies are "too close" (Equation 37). Therefore, the DFT forecast can not succeed.

The analyses of all seven complex time series indicate that the relative accuracy of amplitude and period estimates are lower than the relative accuracy of signal minimum and maximum epoch estimates (Tables 1 - 7). This statistical effect is the same when DCM is applied to the time series of any arbitrary phenomenon. This effect would, for example, explain why our solar cycle amplitude forecasts are less accurate than our solar cycle minimum and maximum epoch forecasts (Jetsu 2025). DCM detects the correct simulated period values, but DFT detects less accurate period values, which are not always correct. The leakage of DFT spectral power (Kay and Marple 1981; Ghaderpour et al. 2021) would explain why correct signal periods have not been detected earlier from the sunspot data.

The DCM analysis results can be double-checked. We use the Fisher-test to identify the best DCM model from all tested DCM models (Equation 48). The correct model may not be among the tested models. Hence, the best model is not necessarily the correct one. This best model must be correct if it passes the Forecast-test (Equation 53). This Fisher- and Forecast-test combination double-checks the DCM analysis results. The perfect result is that the best model is the correct model. Therefore, the Forecast-test also prevents overfitting.

4 Discussion

DCM proceeds through two stages. It computes the R (Equation 14) or χ^2 (Equation 15) values for a massive number of LS fits (Equation 50). The Gauß-Markov theorem ensures that the DCM model having the lowest R or χ^2 is inevitably *always* found, and the result for the non-linear iteration is unique (Equation 22). This first stage devours CPU. The second fast stage, the Fisher-test, then reveals the best DCM model alternative (Equation 48). Practical time series analysis applications, like digital signal processing, require fast computational algorithms (Kay and Marple 1981). Long computation time is the main AL constraining DCM. For example, Jetsu (2025) computed about 4.5 million LS fits to search for four signals in the monthly sunspot data ($n = 3287, K_1 = 4, K_2 = 1, K_3 = 0, n_{\rm L} = 100, n_{\rm S} = 30, n_{\rm B} = 20$). This required several months of CPU. The parallel Python code computations took a few days. Nevertheless, the CPU used is beside the point if a well-posed computational solution can be found to any challenging scientific ill-posed problem.

The DCM computation time AL constraint is amply compensated in real observations because there is no need to wait for the repetition of the signal(-s). The WD-effect is the spearhead of DCM because the sample window (ΔT) can be infinitesimally short. Fast accurate observations are the best approach. If the noise ($\sigma \equiv$ SN) can not be eliminated it is always possible to increase the sample size (n).

It is time to summarise why the ALs of other frequency-domain parametric time series analysis methods (Section 1) do not constrain DCM.

1. Data errors (level of noise) are unknown.

   DCM can solve both alternatives: errors σ_i known or unknown (Equations 16, 17, 46 and 47). All DCM model parameter solutions in Tables 1-7 converge to the correct values when the simulated data n and/or SN increase (σ_i decrease). Due to this WD-effect, sufficient noise reduction always leads to the correct model detection.

2. Data error information is not utilised.

   DCM utilises this information. If the errors σ_i are known, DCM performs weighted LS fits which minimise the model χ^2 for every tested frequency combination (Equation 15). This gives the DCM test statistic z (Equation 17). The Fisher-test utilises the χ^2 values of all alternative tested DCM models to identify the best DCM model (Equation 47).

3. Data must be evenly spaced.

   DCM performance is independent of data spacing. Even or uneven spacing is irrelevant for all n_{Lin} LS fits (Equation 50). However, we do admit that long gaps can mislead even these LS fits.

4. Model parameter errors are unknown.

   DCM gives error estimates for the model parameters of Equations 23-29. For non-linear models, such as the DCM model, the analytical solution for the model parameter errors is a highly complex effort (Furlan and Mortarino 2020). DCM solves these error estimates using the computational statistical bootstrap technique (Equation 31).

5. Model and forecast errors are unknown.

   The computational bootstrap technique (Equation 31) gives n_B estimates for $\boldsymbol{\beta}_J$, $t_{\mathrm{mid,J}}$ and ΔT_J, where $J = 1, 2, ...n_B$. These estimates give n_B values for $g(t)$ at any time t inside and outside the sample window ΔT. The standard deviation of these n_B values gives the error limits $\sigma_{g(t)}$ in Equation 30. The n_B estimates for $\boldsymbol{\beta}_J$, $t_{\mathrm{mid,J}}$ and ΔT_J also give the error limits for the functions $h(t), h_i(t)$ and $p(t)$ in the DCM model $g(t)$.

6. Sample window is shorter than signal period(s).

   DCM solves this problem. The sample window is shorter than the period(s) in Model 1, 2, 5 and 7 time series simulations. Regardless of this, all model parameter estimates converge to correct values when n and/or SN increase (Tables 1, 2, 5 and 7). Due to this WD-effect, the sample window ΔT value is irrelevant for DCM. It can perceive the future and the past much "earlier" (from shorter ΔT) than has been previously thought, like in Equations 35 and 37.

7. Presence and shape of trend are unknown.

   DCM can test any K_1, K_2 and K_3 model combination. The Fisher-test reveals the combination of the best DCM model. The Forecast-test can double-check this result. These tests give the correct K_3 trend order. Due to the WD-effect, the polynomial trend coefficients M_k converge to correct values in all simulated time series (Sections 3.1-3.7). The trend is absent (stationary time series) if $K_3 = -1 \equiv p(t) = 0$ or $K_3 = 0 \equiv p(t) = M_0 = \text{constant}$ (Equations 4 and 5).

8. Sample window causes leakage.

   Leakage does not constrain DCM because the sample window ΔT has no effect on the LS fits (Equation 50: n_{Lin}). All frequencies converge to exactly correct values when the n and/or SN values of simulated time series increase (Tables 1-7). This confirms the absence of leakage. Due to the WD-effect, the performance of DCM does not depend on the sample window ΔT.

9. Leakage weakens frequency resolution.

   There is no leakage because the sample window is irrelevant (AL8). The DCM frequency resolution is limited only by the sample size n and the data accuracy σ. The WD-effect ensures that all frequency detections for the simulated time series converge to exactly correct values (Tables 1-7).

10. Signal shapes are not pure sines.

The DCM always finds the correct K_1, K_2 and K_3 values (AL7). The K_2 value determines the signal shape. The $K_2 = 1$ pure sine shape is the simplest. Higher K_2 values allow the DCM modelling of more complex shapes. We demonstrate the $K_2 = 2$ double wave signal detections in Sections 3.6 and 3.7.

11. Number of signals is unknown.

    The correct K_1, K_2 and K_3 values can always be found (AL7). The K_1 value determines the number of signals. For all simulated time series, the detected signal frequencies and amplitudes converge to exactly correct values (Sections 3.1-3.7). Furthermore, Jetsu (2020) showed that DCM models having too few or too many signals are often unstable ("UM").

12. Correct model alternative is unknown.

    The values of K_1, K_2 and K_3 determine the correct model (AL7). The $(n - \eta_2 - 1)/(\eta_2 - \eta_1)$ penalty terms prevent overfitting when the Fisher-test is used to identify the best DCM model from all tested DCM models (Equations 46 and 47). This best model is also the correct model if it passes the Forecast-test (Equation 53). The WD-effect ensures that better data can inevitably reveal the correct model.

13. Signal significances are unknown.

    The Fisher-test gives the $Q_{\rm F}$ critical levels (signal significances) for all detected signals (Sections 3.9 and 3.10). This $Q_{\rm F}$ is the probability of falsely rejecting the H_0 hypothesis when it is in fact true. Thus, the $Q_{\rm F}$ value represents the probability of false signal detection. This detection is absolutely certain in the $Q_{\rm F} < 10^{-16}$ cases. Our preassigned significance level for signal detection is $\gamma = 0.001$ (Equation 48).

14. Model solution is ill-posed.

    The *analytical* solution for the non-linear DCM model is ill-posed (Equation 1). We show that there exists a well-posed *computational* solution (Section 3.11). The WD-effect ensures that this solution can always be found by increasing the sample size (n) and/or the signal to noise ratio (SN), regardless of the time series complexity (K_1, K_2 and K_3 combination). The main DCM constraint is that the solutions for more complex models require more computation time when the number of signals increases (Equation 50).

15. Complex non-linear model forecasts fail.

    Our computational solution for the non-linear DCM model is well-posed (Section 3.11). The forecast of this well-posed solution can be double-checked (Section 3.12). First, the Fisher-test identifies the best DCM model from all tested DCM models (Section 3.9). Then, the Forecast-test reveals if this best DCM model is also the correct one (Equation 53). Again, the WD-effect ensures that the correct model and forecast can be found for any complex time series (K_1, K_2 and K_3 combination) when n and/or SN increase. We conclude that the DCM rises to meet the challenge of "forecasting the evolution of complex systems" (Cheng et al. 2015).

The DCM turns things upside down. There would be no need for time series analysis, if the correct frequencies were already *known*. The tested frequencies are already *known*. DCM does not search for *unknown* frequencies, it just tests *known* frequencies. DCM tests all possible frequency combinations. The LS fit for every frequency combination gives unique R and χ^2 values. The best frequency combination minimises R or χ^2. This gives the unique initial free parameter $\boldsymbol{\beta}$ values for the non-linear iteration. The frequencies are never *unknown* in this process. The Gauss-Markov theorem, the well-posed computational model solution and the revolutionary WD-effect make absolutely sure that DCM succeeds for any number of signals (K_1), all signal shapes (K_2) and every polynomial trend (K_3). The Fisher-test and the Forecast-test can double-check that the DCM model solution is correct. The DCM analysis of intensive, large and accurate time series can see through time: a glimpse of the future and the past.

Our DCM is a remarkable method because it outperforms the DFT. Gauß (1777–1855) developed an early form of the Fast Fourier Transform (FFT) algorithm for astronomical purposes, but

did he not publish it. Perhaps our DCM should be re-named the Slow Gauß Transform (SGT).

5 Appendix

In this appendix, we show that our DCM can model and forecast the El Niño phenomenon between the years 1870 and 2024. This result may be the "holy grail" of climatology. The solar forcing most probably causes the detected 5.850 ± 0.085, 12.82 ± 0.40 and 19.30 ± 0.83 year signals. These strict periodicities indicate that the solar cycle is caused by the planets. Our results confirm that the DCM is an ideal time series analysis method for detecting an unknown number of signals superimposed on an unkown trend.

The El Niño phenomenon is considered unpredictable (Thirumalai et al. 2024; Lu et al. 2025; Cai et al. 2014; Hu et al. 2024; Liu et al. 2023; Ludescher et al. 2013; Timmermann et al. 2018; Liang et al. 2021). Our analysed data are the yearly means in the The Seasonal Mean Nino 4 HadISST1.1 (NOAA PSL) sample.[5] There were no values for the last four months of 2025. The full data contains $n = 155$ yearly means between 1870 and 2024.

5.1 Trend

We detect significant signals only from the weighted data. Therefore, the trend and the forecasts are solved from the weighted data. The weighted DCM analysis results for all data are given in Table S1. For one, two, three and four signals, the best trend alternative is the linear trend $K_1 = 1$. We use this $K_1 = 1$ trend in our DCM analysis of El Niño data. For the $\mathcal{M}$=3 model, the positive linear trend is $2M_1 = 0.56 \pm 0.20$ C^{o}during $\Delta T = 154$ years (Equation 5), like in all $\mathcal{M}$=1-4 models.

5.2 Forecasting data

We use the first half of all data ($n = 78$) to forecast the second half ($n = 77$). The results of weighted DCM analysis are given in Table S2. The periods of the three strongest signals are 5.850 ± 0.085, 12.82 ± 0.40 and 19.30 ± 0.83 years (Model $\mathcal{M}$=3). The forecasts of the two and three signals models, open circles, are shown in Figures S1 and S2. For the second half of the data, the forecast χ^2 values for the one, two, three and four signals models are 421, 336, 370 and 428, respectively. Hence, the two signal $\mathcal{M}$=2 model gives the best forecast.

5.3 All data

The DCM analysis results for all weighted data are given in Table S3. The three signals $\mathcal{M}$=3 model is the best. The four signal model is unstable (UM). The three and four signal models are shown in Figures S3 and S4. For both alternatives, the predicted values, open circles, after 2024 are given in Table S4. Both alternatives give essentially the same forecast.

For the sake of consistency, we also give the results of non-weighted analysis of all data (Table S5). We detect the same signals as from non-weighted data, but the signal significances Q_{F} are lower. This confirms that the data error information is crucial.

5.4 Main result

Our DCM can model and forecast the El Niño phenomenon. The solar forcing is the only possible explanation for these results. Within their error limits, the detected three El Niño signals could represent the Hale cycle (Hale et al. 1919), the solar cycle(Schwabe 1844; Wolf 1852) and half the solar cycle. These strict El Niño periodicities indicate that the planets could cause a stationary and deterministic solar cycle (Jetsu 2025). A stochastic and non-stationary solar cycle (Charbonneau 2020; Usoskin 2023) could not cause the detected regular El Niño periodicities.

[5] The data from https://psl.noaa.gov/data/timeseries/month/data/nino4.long.anom.data were retrieved on January 29th, 2026.

Acknowledgements

This work has made use of NASA's Astrophysics Data System (ADS) services.

Author contributions statement

L.J. peformed this research and wrote the manuscript.

Consent for publication

The author grants permission for publication.

Additional information

Supplementary information The online version contains supplementary material available.

Data availability

Main manuscript analysis: All data files and all DCM control files are stored to https://zenodo.org/uploads/17018676.

Appendix analysis: All files, DCM Python code, DCM control files and Python code are stored to the https://zenodo.org/records/18595333.

Declarations

Competing interests

The author declares no competing interests.

Correspondence and requests for materials should be addressed to L.J.

Funding No funding was received.

Ethical approval and consent to participate Not applicable.

Supplementary material

The $\mathcal{M}=$ 1 analysis in Table S1 can be repeated using these two commands

```
cp longC14K110.dat dcm.dat
python dcm.py
```

The whole El Niño analysis can be repeated in similar fashion.

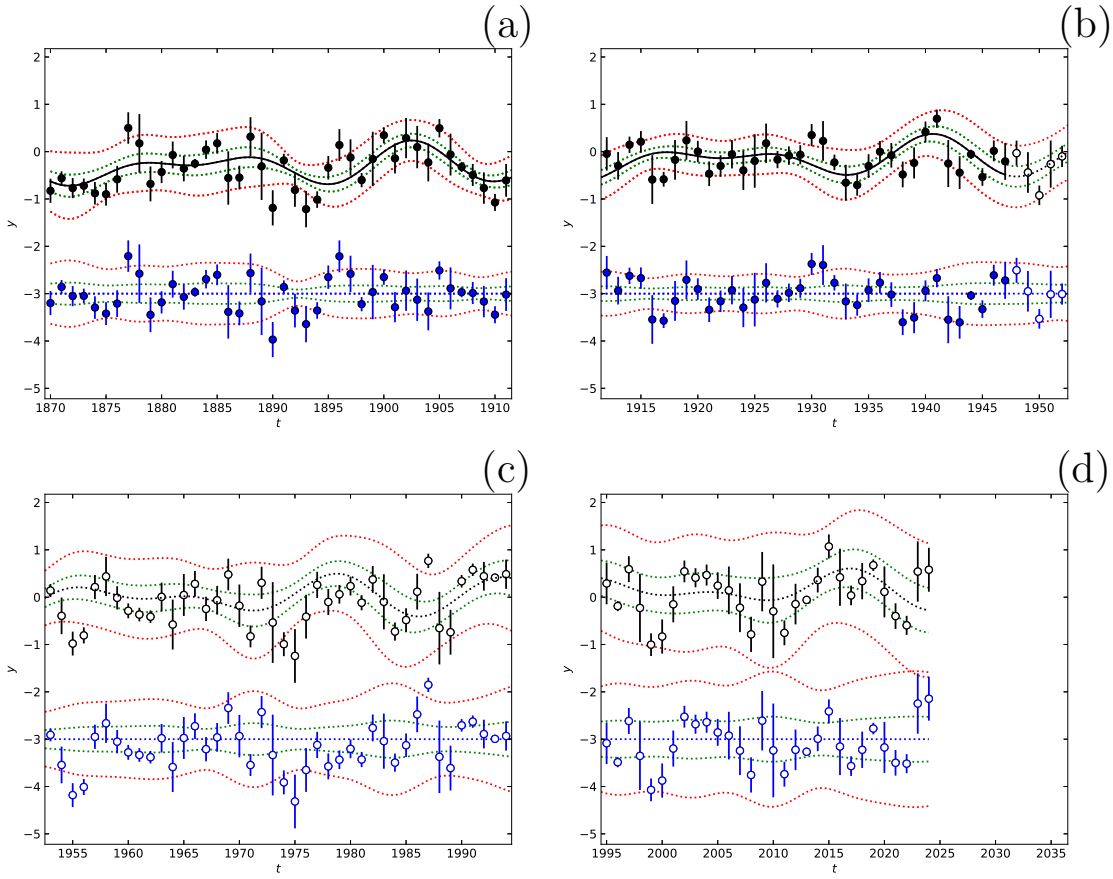

Fig. S1 First half forecast (Table S2: $\mathcal{M}$=2): Forecasting data (closed black circles) and forecasted data (open black circles). Green and red dottted lines denote one and three sigma model $g(t)$ errors. Residuals (closed and open blue circles) are offset to level of -3.

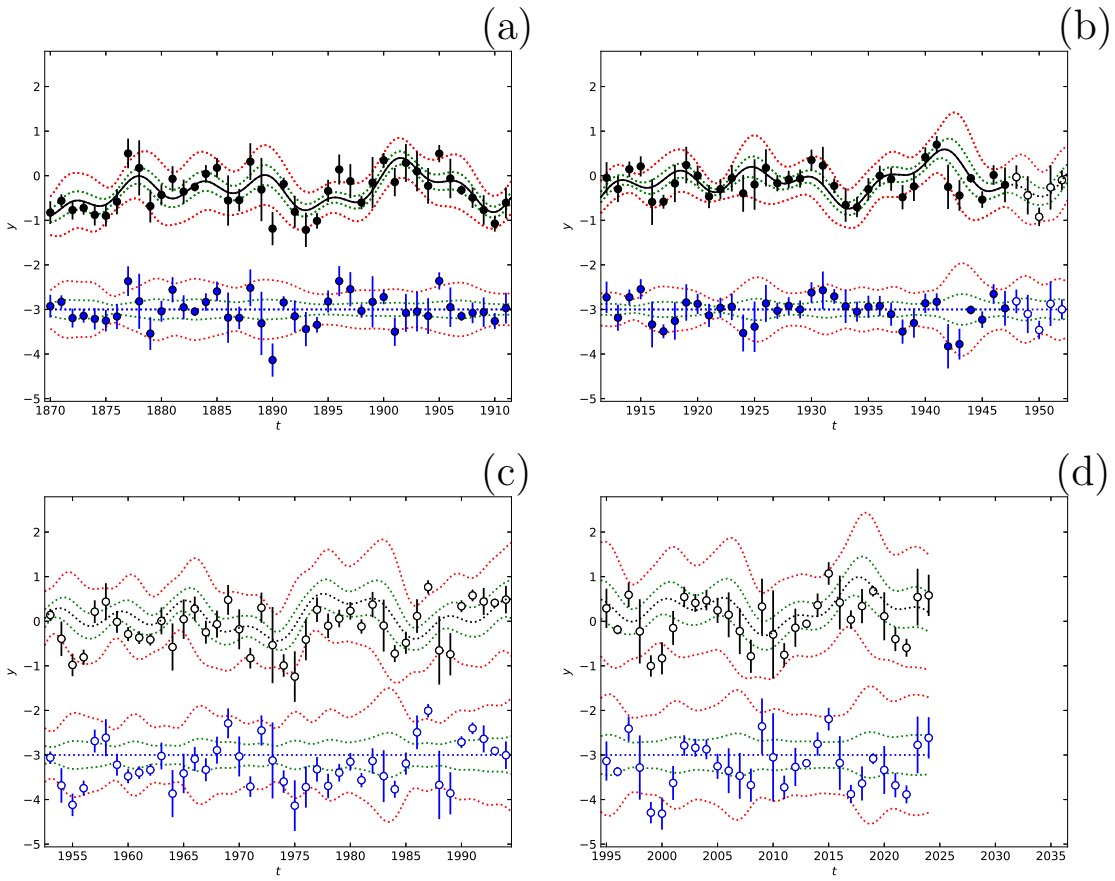

Fig. S2 First half forecast (Table S2: $\mathcal{M}$=3). Otherwise as in Figure S1.

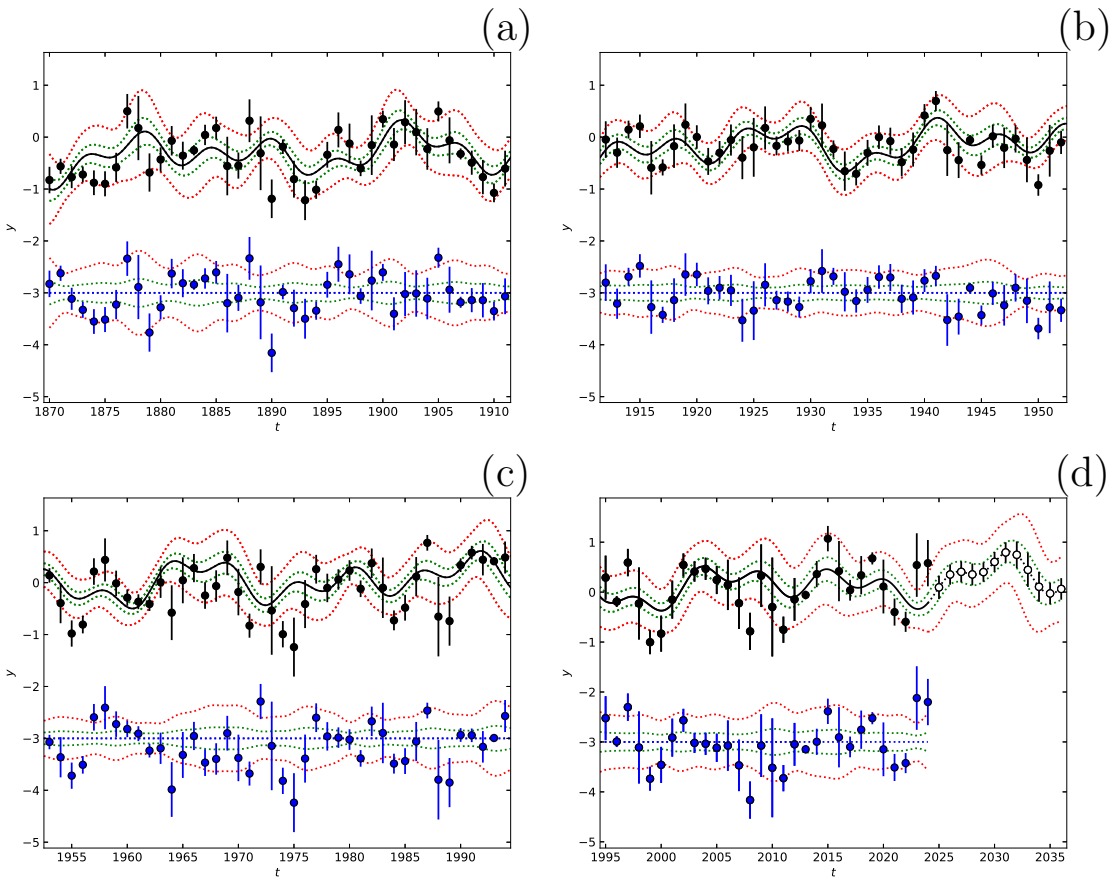

Fig. S3 All data model $\mathcal{M}$=3 in Table S3. Otherwise as in Figure S1.

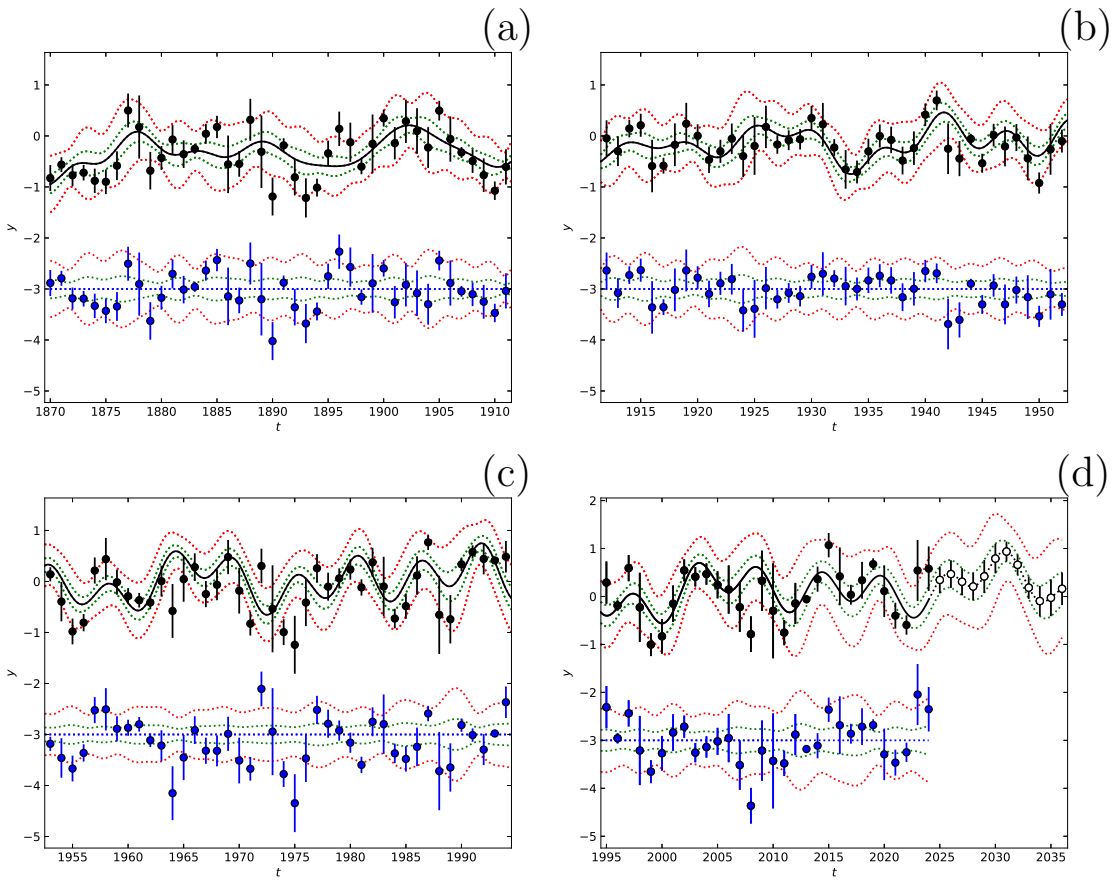

Fig. S4 All data model $\mathcal{M}$=4 in Table S3. Otherwise as in Figure S1.

Table S1 Trend in all data. Weighted DCM analysis performed between $P_{\min} = 3$ and $P_{\max} = 100$. (1) $\mathcal{M}$model, (2) η free parameters and χ^2 of model, (3) First signal: P_1 period, A_1 peak to peak amplitude and $t_{\min,1}$ primary minimum epoch. (4-6) Next signals. (7-8) Fisher-test: F_χ test statistic and $Q_{\rm F}$ critical level. (9) Control file. Arrows point to best model in Fisher-test comparison.

		Period analysis				Fisher-test		
		Data: Original weighted data ($n = 155, \Delta T = 154$: `Clong.dat`)						
$\mathcal{M}$	η (-)	P_1 (y)	P_2 (y)	P_3 (y)	P_4 (y)	$\mathcal{M}$=2	$\mathcal{M}$=3	Control file
	χ^2 (-)	A_1 (C^o)	A_2 (C^o)	A_3 (C^o)	A_4 (C^o)	F_χ (-)	F_χ (-)	
		$t_{\min,1}$ (y)	$t_{\min,1}$ (y)	$t_{\min,1}$ (y)	$t_{\min,1}$ (y)	Q_F (-)	Q_F (-)	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
One signal								
$\mathcal{M}$=1		12.67 ± 0.33	-	-	-	↑	↑	
$\mathcal{M}_{1,1,0,\chi^2}$	4	0.62 ± 0.15	-	-	-	26.9	14.0	`longC14K111.dat`
	516	1871.2 ± 1.3	-	-	-	6.7×10^7	2.7×10^{-6}	
$\mathcal{M}$=2		12.65 ± 0.20	-	-	-	-	←	
$\mathcal{M}_{1,1,1,\chi^2}$	5	0.51 ± 0.14	-	-	-	-	1.02	`longC14K111.dat`
	437	1871.3 ± 1.3	-	-	-	-	0.31	
$\mathcal{M}$=3		12.65 ± 0.43			-	-	-	
$\mathcal{M}_{1,1,2,\chi^2}$	6	0.52 ± 0.12			-	-	-	`longC41K112.dat`
	434	1871.4 ± 1.1			-	-	-	
Two signals								
						$\mathcal{M}$=5	$\mathcal{M}$=6	
$\mathcal{M}$=4		12.70 ± 0.17	26.8 ± 2.0	-	-	↑	↑	
$\mathcal{M}_{2,1,0,\chi^2}$	7	0.62 ± 0.12	0.39 ± 0.16	-	-	27.4	13.6	`longC14K210.dat`
	456	1870.8 ± 1.0	1870.3 ± 3.4	-	-	5.6×10^{-7}	3.8×10^{-6}	
$\mathcal{M}$=5		5.677 ± 0.089	12.64 ± 0.31	-	-	-	←	
$\mathcal{M}_{2,1,1,\chi^2}$	8	0.393 ± 0.086	0.47 ± 0.13	-	-	-	0	`longC14K211.dat`
	384	$1875.65 \pm .56$	1871.3 ± 1.2	-	-	-	1	
$\mathcal{M}$=6		5.677 ± 0.077	12.64 ± 0.42		-	-	-	
$\mathcal{M}_{2,1,2,\chi^2}$	9	0.388 ± 0.093	0.47 ± 0.12		-	-	-	`longC41K212.dat`
	384	1875.66 ± 0.76	1871.4 ± 1.2		-	-	-	
Three signals								
						$\mathcal{M}$=8	$\mathcal{M}$=9	
$\mathcal{M}$=7		5.619 ± 0.091	12.74 ± 0.48	21.83 ± 0.81	-	↑	↑	
$\mathcal{M}_{3,1,0,\chi^2}$	10	0.41 ± 0.13	0.52 ± 0.11	0.40 ± 0.12	-	29.2	15.0	`longC14K310.dat`
	407	1871.04 ± 0.77	1870.7 ± 1.7	1870.0 ± 3.2	-	2.6×10^{-7}	1.2×10^{-6}	
$\mathcal{M}$=8		5.662 ± 0.077	12.77 ± 0.12	21.3 ± 1.5	-	-	←	
$\mathcal{M}_{3,1,1,\chi^2}$	11	0.40 ± 0.13	0.47 ± 0.10	0.37 ± 0.12	-	-	0.82	`longC14K311.dat`
	338	1870.3 ± 0.62	1870.4 ± 0.87	1870.8 ± 3.2	-	-	0.36	
$\mathcal{M}$=9		5.661 ± 0.073	12.76 ± 0.22	21.2 ± 1.2	-	-	-	
$\mathcal{M}_{3,1,2,\chi^2}$	12	0.40 ± 0.11	0.49 ± 0.10	0.39 ± 0.12	-	-	-	`longC41K312.dat`
	336	1870.3 ± 0.74	1870.6 ± 1.2	1871.1 ± 2.8	-	-	-	
Four signals								
						$\mathcal{M}$=11	$\mathcal{M}$=12	
$\mathcal{M}$=10		5.597 ± 0.066	12.77 ± 0.19	19.4 ± 1.6 IF	21.5 ± 2.1 IF	↑	↑	
$\mathcal{M}_{4,1,0,\chi^2}$	13	0.44 ± 0.10	0.57 ± 0.11	0.4 ± 1.3 AD	0.4 ± 1.4 *AD*	28.1	14.2	`longC14K410.dat`
UM	362	1871.56 ± 0.57	1870.3 ± 1.2	1872.8 ± 3.7	1871.8 ± 3.3	4.5×10^{-7}	2.5×10^{-6}	
$\mathcal{M}$=11		5.496 ± 0.093 IF	5.66 ± 0.13 IF	12.76 ± 0.15	21.1 ± 1.5	-	←	
$\mathcal{M}_{4,1,1,\chi^2}$	14	0.3 ± 1.8 AD	0.4 ± 1.8 AD	0.49 ± 0.13	0.39 ± 0.11	-	0.45	`longC14K411.dat`
UM	300	1873.8 ± 1.2	1870.2 ± 1.2	1870.60 ± 0.97	1871.3 ± 2.1	-	0.50	
$\mathcal{M}$=12		5.497 ± 0.087 IF	0.566 ± 0.062 IF	12.75 ± 0.17	21.0 ± 1.0	-	-	
$\mathcal{M}_{4,1,2,\chi^2}$	15	0.34 ± 0.94 AD	0.42 ± 0.90 AD	0.50 ± 0.11	0.398 ± 0.084	-	-	`longC41K412.dat`
UM	299	1873.8 ± 1.0	1870.18 ± 0.99	1870.6 ± 0.97	1871.4 ± 2.2	-	-	

Table S2 Periods in first half data. Otherwise as in Table S1.

		Period analysis				Fisher-test			
		Data: Original weighted data ($n = 78, \Delta T = 77$: `Chalf.dat`)							
$\mathcal{M}$	η (-)	P_1 (y)	P_2 (y)	P_3 (y)	P_4 (y)	$\mathcal{M}$=2	$\mathcal{M}$=3	$\mathcal{M}$=4	Control file
	χ^2 (-)	A_1 (C^o)	A_2 (C^o)	A_3 (C^o)	A_4 (C^o)	F_χ (-)	F_χ (-)	F_χ(-)	
		$t_{\min,1}$ (y)	$t_{\min,1}$ (y)	$t_{\min,1}$ (y)	$t_{\min,1}$ (y)	Q_F (-)	Q_F (-)	Q_F (-)	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
One signal									
$\mathcal{M}$=1		19.1 ± 1.2	-	-	-	←	↑	↑	
$\mathcal{M}_{1,1,1,\chi^2}$	5	0.44 ± 0.12	-	-	-	5.62	5.95	5.78	`halfC14K111.dat`
	168	1873.6 ± 2.0	-	-	-	0.0016	5.0×10^{-5}	7.5×10^{-6}	
Two signals									
$\mathcal{M}$=2		12.82 ± 0.54	18.78 ± 0.65	-	-	-	←	↑	
$\mathcal{M}_{2,1,1,\chi^2}$	8	0.44 ± 0.12	0.53 ± 0.13	-	-	-	5.3	4.91	`halfC14K211.dat`
	135	1870.4 ± 1.4	1874.3 ± 1.9	-	-	-	0.0025	0.00034	
Three signals									
$\mathcal{M}$=3		5.580 ± 0.085	12.82 ± 0.40	19.32 ± 0.83	-	-	-	←	
$\mathcal{M}_{3,1,1,\chi^2}$	11	0.42 ± 0.14	0.488 ± 0.087	0.499 ± 0.096	-	-	-	3.88	`halfC14K311.dat`
	109	1875.01 ± 0.51	1870.2 ± 1.4	1873.5 ± 2.3	-	-	-	0.012	
Four signals									
$\mathcal{M}$=4		3.437 ± 0.052 IF	5.70 ± 0.13IF	$12.71 \pm x0.17$	21.2 ± 1.2	-	-	-	
$\mathcal{M}_{4,1,1,\chi^2}$	14	0.30 ± 0.12 AD	0.40 ± 0.31 AD	0.46 ± 0.13	0.440 ± 0.096	-	-	-	`halfC14K411.dat`
	92	1872.51 ± 0.66	1875.6 ± 1.2	1870.8 ± 1.1	1870.9 ± 2.9	-	-	-	

Table S3 Periods in all weighted data. Otherwise as in Table S1.

	Period analysis					Fisher-test			
	Data: Original weighted data ($n = 155, \Delta T = 154$: `Clong.dat`)								
$\mathcal{M}$		P_1 (y)	P_2 (y)	P_3 (y)	P_4 (y)	$\mathcal{M}$=2	$\mathcal{M}$=3	$\mathcal{M}$=4	
	η (-)	A_1 (C^o)	A_2 (C^o)	A_3 (C^o)	A_4 (C^o)	F_χ (-)	F_χ (-)	F_χ(-)	Control file
	χ^2 (-)	$t_{\min,1}$ (y)	$t_{\min,1}$ (y)	$t_{\min,1}$ (y)	$t_{\min,1}$ (y)	Q_F (-)	Q_F (-)	Q_F (-)	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
One signal									
$\mathcal{M}$=1		12.65 ± 0.20	-	-	-	↑	↑	↑	
$\mathcal{M}_{1,1,1,\chi^2}$	5	0.51 ± 0.14	-	-	-	12.4	6.98	7.10	`longC14K111.dat`
	437	1871.3 ± 1.3	-	-	-	2.7×10^{-7}	1.6×10^{-6}	1.9×10^{-8}	
Two signals									
$\mathcal{M}$=2		5.677 ± 0.089	12.64 ± 0.31	-	-	-	↑	↑	
$\mathcal{M}_{2,1,1,\chi^2}$	8	0.393 ± 0.086	0.47 ± 0.13	-	-	-	6.48	6,53	`longC14K211.dat`
	384	1875.65 ± 0.56	1871.3 ± 1.2	-	-	-	0.00038	4.1×10^{-6}	
Three signals									
$\mathcal{M}$=3		5.662 ± 0.077	12.78 ± 0.12	21.3 ± 1.5	-	-	-	↑	
$\mathcal{M}_{3,1,1,\chi^2}$	11	0.40 ± 0.13	0.47 ± 0.10	0.37 ± 0.12	-	-	-	5.91	`longC14K311.dat`
	338	1870.27 ± 0.62	1870.36 ± 0.88	1870.8 ± 3.3	-	-	-	0.00079	
Four signals									
$\mathcal{M}$=4		5.496 ± 0.093 IF	5.66 ± 0.13 IF	12.76 ± 0.15	21.09 ± 1.5	-	-	-	
$\mathcal{M}_{4,1,1,\chi^2}$	14	0.3 ± 1.8 AD	0.4 ± 1.8 AD	0.491 ± 0.13	0.39 ± 0.11	-	-	-	`longC14K411.dat`
UM	300	1873.8 ± 1.2	1870.2 ± 1.1	1870.6 ± 0.97	1871.3 ± 2.	-	-	-	

Table S4 Forecasts (Figures S3 and S4: open circles).

	Figure S3		Figure S4	
t (y)	y (C^o)	σ (C^o)	y (C^o)	σ (C^o)
2025	0.09	0.24	0.35	0.28
2026	0.35	0.20	0.47	0.29
2027	0.41	0.21	0.31	0.28
2028	0.35	0.22	0.21	0.28
2029	0.40	0.19	0.42	0.34
2030	0.60	0.20	0.79	0.31
2031	0.79	0.20	0.94	0.21
2032	0.75	0.27	0.66	0.21
2033	0.45	0.34	0.18	0.23
2034	0.11	0.30	-0.10	0.31
2035	-0.03	0.21	-0.03	0.38
2036	0.06	0.22	0.16	0.34

Table S5 Periods in all non-weighted data. Otherwise as in Table S1.

		Period analysis				Fisher-test			
		Data: Original non-weighted data ($n = 155, \Delta T = 154$: `Rlong.dat`)							
$\mathcal{M}$		P_1 (y)	P_2 (y)	P_3 (y)	P_4 (y)	$\mathcal{M}$=2	$\mathcal{M}$=3	$\mathcal{M}$=4	
	η (-)	A_1 (-)	A_2 (-)	A_3 (-)	A_4 (-)	F_R (-)	F_R (-)	F_R(-)	Control file
	R (-)	$t_{\mathrm{min},1}$ (y)	$t_{\mathrm{min},1}$ (y)	$t_{\mathrm{min},1}$ (y)	$t_{\mathrm{min},1}$ (y)	Q_F (-)	Q_F (-)	Q_F (-)	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
One signal									
$\mathcal{M}$=1		12.71 ± 0.12	-	-	-	←	←	←	
$\mathcal{M}_{1,1,1,R}$	5	0.389 ± 0.074	-	-	-	3.67	3.77	1.11	`longR14K110.dat`
	27.1	1871.19 ± 0.86	-	-	-	0.013	0.0016	0.36	
Two signals									
$\mathcal{M}$=2		5.651 ± 0.030	12.69 ± 0.15	-	-	-	←	←	
$\mathcal{M}_{2,1,1,R}$	8	0.319 ± 0.064	0.384 ± 0.088	-	-	-	3.77	-0.092	`longR14K211.dat`
	25.2	1870.26 ± 0.41	1871.3 ± 1.1	-	-	-	0.0016	1	
Three signals									
$\mathcal{M}$=3		5.650 ± 0.032	9.12 ± 0.12	12.69 ± 0.27	-	-	-	←	
$\mathcal{M}_{3,1,1,R}$	11	0.327 ± 0.084	0.299 ± 0.087	0.381 ± 0.079	-	-	-	-3.50	`longR14K311.dat`
	23.4	1870.28 ± 0.46	1872.5 ± 1.2	1871.4 ± 1.3	-	-	-	1	
Four signals									
$\mathcal{M}$=4		5.496 ± 0.093 IF	5.662 ± 1.3 IF	12.76 ± 0.16	21.1 ± 1.5	-			
$\mathcal{M}_{4,1,1,R}$	14	0.3 ± 1.9 AD	0.4 ± 1.8 AD	$0.491 \pm 0.49 \pm 0.13$	0.39 ± 0.11	-			`longR14K411.dat`
UM	25.3	1873.8 ± 1.2	1870.2 ± 1.2	1870.6 ± 0.97	1871.3 ± 2.1	-			

References

Bard, Y.: Nonlinear Parameter Estimation. Academic Press, New York, NY (1974)

Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods. Springer, New York, NY (2009)

Berger, J.O.: Statistical Decision Theory and Bayesian Analysis. Springer, London (2013)

Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: Time Series Analysis: Forecasting and Control. John Wiley & Sons, Hoboken, New Jersey (2015)

Bretthorst, G.L.: Excerpts from bayesian spectrum analysis and parameter estimation. In: Maximum-Entropy and Bayesian Methods in Science and Engineering: Foundations, pp. 75–145. Springer, London (1988)

Cai, W., Borlace, S., Lengaigne, M., Van Rensch, P., Collins, M., Vecchi, G., Timmermann, A., Santoso, A., McPhaden, M.J., Wu, L., *et al.*: Increasing frequency of extreme el niño events due to greenhouse warming. Nature climate change **4**(2), 111–116 (2014)

Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I.: Stl: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics **6**(1), 3–73 (1990)

Charbonneau, P.: Dynamo models of the solar cycle. Living Reviews in Solar Physics **17**(4), 1–104 (2020) https://doi.org/10.1007/s41116-020-00025-6

Cheng, C., Sa-Ngasoongsong, A., Beyca, O., Le, T., Yang, H., Kong, Z., Bukkapatnam, S.T.: Time series forecasting for nonlinear and non-stationary processes: a review and comparative study. Iie Transactions **47**(10), 1053–1071 (2015)

Cooley, J.W., Tukey, J.W.: An algorithm for the machine calculation of complex fourier series. Mathematics of computation **19**(90), 297–301 (1965)

Draper, N.R., Smith, H.: Applied Regression Analysis. John Wiley & Sons, Inc., Hoboken, New Jersey (1998). https://doi.org/10.1002/9781118625590

Engl, H.W., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Mathematics and Its Applications. Kluwer Academic Publishers Group, Dordrecht (1996)

Efron, B., Tibshirani, R.: Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. Statistical Science **1**, 54 (1986)

Furlan, C., Mortarino, C.: Comparison among simultaneous confidence regions for nonlinear diffusion models. Computational Statistics **35**(4), 1951–1991 (2020)

Fourier, J.B.J.: Théorie Analytique de la Chaleur, (1822)

Gauß, K.F.: Theoria Motvs Corporvm Coelestivm in Sectionibvs Conicis Solem Ambientivm., (1809)

Gauß, C.F.: Theoria Combinationis Observationum Erroribus Minimis Obnoxiae. Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores **1**, 1–94 (1821)

Grushka-Cockayne, Y., Jose, V.R.R., Lichtendahl Jr, K.C.: Ensembles of overfit and overconfident forecasts. Management Science **63**(4), 1110–1130 (2017)

Ghaderpour, E., Pagiatakis, S.D., Hassan, Q.K.: A survey on change detection and time series analysis with applications. Applied Sciences **11**(13) (2021) https://doi.org/10.3390/app11136141

Hadamard, J.: Sur les problèmes aux dérivées partielles et leur signification physique. Princeton University Bulletin **13**, 49–52 (1902)

Hadamard, J.: Lectures on Cauchy's Problem in Linear Partial Differential Equations. Mrs. Hepsa Ely Silliman memorial lectures. Yale University Press, New Haven, Connecticut (1923)

Hamilton, J.D.: Time Series Analysis. Princeton University Press, Princeton (1994). https://doi.

org/10.1515/9780691218632

Horne, J.H., Baliunas, S.L.: A Prescription for Period Analysis of Unevenly Sampled Time Series. *Astrophys. J.* **302**, 757 (1986) https://doi.org/10.1086/164037

Hale, G.E., Ellerman, F., Nicholson, S.B., Joy, A.H.: The Magnetic Polarity of Sun-Spots. *Astrophys. J.* **49**, 153 (1919) https://doi.org/10.1086/142452

Hu, R., Lian, T., Liu, T., Wang, J., Song, X., Chen, H., Chen, D.: Predicting the 2023/24 el niño from a multi-scale and global perspective. Communications Earth & Environment **5**(1), 675 (2024)

Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer Series in Statistics. Springer, New York, NY, USA (2001)

Jetsu, L.: Discrete Chi-square Method for Detecting Many Signals. The Open Journal of Astrophysics **3**(1), 4 (2020) https://doi.org/10.21105/astro.2002.03890 arXiv:2002.03890

Jetsu, L.: Do the planets cause the sunspot cycle? arXiv e-prints, 2311–08317 (2025) https://doi.org/10.48550/arXiv.2311.08317 arXiv:2311.08317. Submitted to Scientific Reports, August 13th, 2025

Jetsu, L., Pelt, J.: Three stage period analysis and complementary methods. *Astron. Astrophys. Suppl.* **139**, 629–643 (1999) https://doi.org/10.1051/aas:1999411

Kay, S.M., Marple, S.L.: Spectrum analysis—a modern perspective. Proceedings of the IEEE **69**(11), 1380–1419 (1981) https://doi.org/10.1109/PROC.1981.12184

Kazemi, M., Rodrigues, P.C.: Robust singular spectrum analysis: comparison between classical and robust approaches for model fit and forecasting. Computational Statistics **40**(6) (2025)

Liu, Y., Donat, M.G., England, M.H., Alexander, L.V., Hirsch, A.L., Delgado-Torres, C.: Enhanced multi-year predictability after el niño and la niña events. Nature communications **14**(1), 6387 (2023)

Lefrancois, P.: Confidence intervals for non-stationary forecast errors: Some empirical results for the series in the m-competition. International Journal of Forecasting **5**(4), 553–557 (1989)

Legendre, A.-M.: Nouvelles méthodes pour la détermination des orbites des comètes. Journal de l'École Polytechnique (1805)

Ludescher, J., Gozolchiani, A., Bogachev, M.I., Bunde, A., Havlin, S., Schellnhuber, H.J.: Improved el niño forecasting by cooperativity detection. Proceedings of the National Academy of Sciences **110**(29), 11742–11745 (2013)

Lehtinen, J., Jetsu, L., Hackman, T., Kajatkari, P., Henry, G.W.: The continuous period search method and its application to the young solar analogue HD 116956. *Astron. Astrophys.* **527**, 136 (2011) https://doi.org/10.1051/0004-6361/201015454 arXiv:1007.4090

Lomb, N.R.: Least-Squares Frequency Analysis of Unequally Spaced Data. *Astrophys. Space Sci.* **39**(2), 447–462 (1976) https://doi.org/10.1007/BF00648343

Lavrentiev, M.M., Romanov, V.G., Shishatskii, S.P.: Ill-posed Problems of Mathematical Physics and Analysis. Translations of Mathematical Monographs, vol. 64. American Mathematical Society, Providence, R.I. (1986)

Lu, Z., Schultze, A., Carré, M., Brierley, C., Hopcroft, P.O., Zhao, D., Zheng, M., Braconnot, P., Yin, Q., Jungclaus, J.H., et al.: Increased frequency of multi-year el niño–southern oscillation events across the holocene. Nature Geoscience, 1–7 (2025)

Liang, X.S., Xu, F., Rong, Y., Zhang, R., Tang, X., Zhang, F.: El niño modoki can be mostly predicted more than 10 years ahead of time. Scientific Reports **11**(1), 17860 (2021)

Makridakis, S., Hibon, M., Lusk, E., Belhadjali,

M.: Confidence intervals: An empirical investigation of the series in the m-competition. International Journal of Forecasting **3**(3-4), 489–508 (1987)

Nerlove, M.: Spectral analysis of seasonal adjustment procedures. Econometrica: Journal of the Econometric Society, 241–286 (1964)

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M.Z., Barrow, D.K., Taieb, S.B., Bergmeir, C., Bessa, R.J., Bijak, J., Boylan, J.E., *et al.*: Forecasting: theory and practice. International Journal of forecasting **38**(3), 705–871 (2022)

Rudin, W.: Principles of Mathematical Analysis. International series in pure and applied mathematics. McGraw-Hill, New York, NY (1976)

Scargle, J.D.: Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.* **263**, 835–853 (1982) https://doi.org/10.1086/160554

Schwabe, M.: Sonnenbeobachtungen im Jahre 1843. Von Herrn Hofrath Schwabe in Dessau. Astronomische Nachrichten **21**, 233 (1844)

Shumway, R.H., Stoffer, D.S.: Time Series Analysis and Its Applications: with R Examples. Springer, New York, NY (2006)

Su, E.C.-Y., Wu, H.-M.: Dimension reduction and visualization of multiple time series data: a symbolic data analysis approach. Computational Statistics **39**(4), 1937–1969 (2024)

Tikhonov, A.N., Arsenin, V.Y.: Solutions of Ill-posed Problems, p. 258. V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York (1977). Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics

Timmermann, A., An, S.-I., Kug, J.-S., Jin, F.-F., Cai, W., Capotondi, A., Cobb, K.M., Lengaigne, M., McPhaden, M.J., Stuecker, M.F., *et al.*: El niño–southern oscillation complexity. Nature **559**(7715), 535–545 (2018)

Tsay, R.S., Chen, R.: Nonlinear Time Series Analysis. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, New Jersey (2018)

Thirumalai, K., DiNezio, P.N., Partin, J.W., Liu, D., Costa, K., Jacobel, A.: Future increase in extreme el niño supported by past glacial changes. Nature **634**(8033), 374–380 (2024)

Tong, H.: Non Linear Time Series: A Dynamical System Approach. Clarendon Press, Oxford, England (1990)

Usoskin, I.G.: A history of solar activity over millennia. Living Reviews in Solar Physics **20**(1), 2 (2023) https://doi.org/10.1007/s41116-023-00036-z

Vogel, C.R.: Computational Methods for Inverse Problems. SIAM Classics in Applied Mathematics, vol. 29. SIAM, Philadelphia, Pennsylvania (2002)

Wolf, R.: Bericht über neue Untersuchungen über die Periode der Sonnenflecken und ihrer Bedeutung von Herrn Prof. Wolf. Astronomische Nachrichten **35**(25), 369 (1852) https://doi.org/10.1002/asna.18530352504

Wooldridge, J.M.: Econometric Analysis of Cross Section and Panel Data, 2nd edn. The MIT Press, Cambridge, MA, USA (2010)