Name:	
Student ID-number:	
Computer Username	
E-mail address	
Study field: Ast..,Phys..?	
Courses:	e.g. basic courses in astronomy and physics
Programming experience:	e.g. IDL, Latex otherwise inexperienced
Questions/ Wishes:	e.g. Is this a too difficult course for me? Could I stay home and get 5 pts?
Anonymous .... clip here	.... clip here ....
Criticism/ Suggestions:	e.g. This course was: incomprehensible, useless boring, and above all badly organised. My contribution was overlooked.

1. Time series of unevenly spaced data

1.1. Backgroud

- Lauri Jetsu
    - ⊙ Ph.D., 1994
    - ⊙ Docent, 1996
    - ⊙ Director of Observatory, 2001–2009
    - ⊙ University Lecturer, 2010
- Aim: To learn, program and apply the Three Stage Period Analysis and Complementary Methods
    - ⊙ TSPA=Three Stage Period Analysis
    - ⊙ PAPER I: Jetsu, L. & Pelt, J. 1999, A&AS **139**, 629
- Homepage of the Course: `http://www.helsinki.fi/~jetsu/time1/time1.html`
- Only English
    - ⊙ Ask about words when necessary.
- We shall use `emacs` to edit our files/programs.
- Our programming language will be IDL (Interactive Data Language)
- If IDL help (`?`) does not work, edit the line `export IDL_HELP_BROWSER=firefox` into file `.bashrc` in `sky.astro.helsinki.fi`
- Evaluation
    - ⊙ No exam(s), only assignments
    - ⊙ If many participants, then division into groups (fill the form).
    - ⊙ Assignments during lectures and also as homework.
    - ⊙ No competion: Groups can/will help each other
    - ⊙ You explain topics to each other, and to me.
    - ⊙ Do not get embarrassed, just say: "I don't know."
    - ⊙ Do not memorize, try to understand instead or ask.
    - ⊙ I/we omit details on purpose, e.g. d/dx(sin x)=?
- Lecture: New concept(-s) explained ⇒ assignment(-s).
    - ⊙ Time limit for assignments expires at the next lecture. You will show and explain your solutions of personal assignments to others. Solutions for all assignments always explained to everybody "until it gets boring".

2. Real phenomena

- Some application examples: Sunspots and starspots

  ⊙ Visibility of sunspots changes as the Sun rotates ⇒ Rotational modulation of brightness, but only about ±0.1% in flux

  ⊙ Larger areas ⇒ Larger amplitudes

  ⊙ More sunpots ⇒ More minima and maxima

  ⊙ Differential rotation ⇒ Different periods at different latitudes

  ⊙ Sunspots appear and disappear ⇒ Minima appear and disappear, Periods change, ...

  ⊙ Sunspot cycle of 11 years ⇒ Mean, Amplitude and Period changes

  ⊙ Cycle irregular ⇒ from 9 to 14 years

  ⊙ Sunspots can disappear for decades (Maunder minimum)

  ⊙ Butterfly diagram ⇒ Regular latitude changes during sunspot cycle

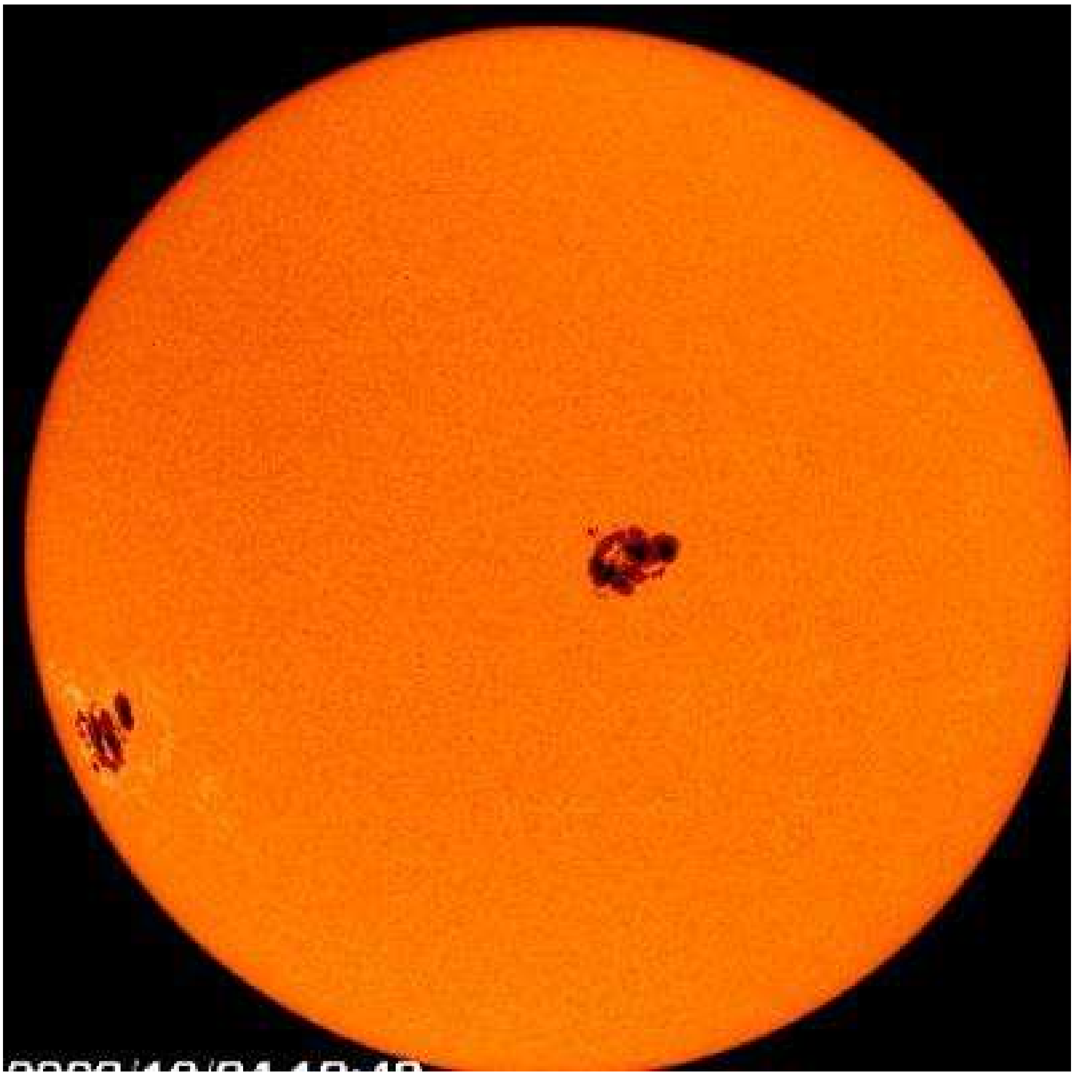

Fig. 1. Sunspots observed with the SOHO-satellite.

- Solar variations are small, about 0.1% in flux
- Stellar variations caused by starspots can be large, even 50% in flux
    - ⊙ X-axis is phase ≡ longitude
    - ⊙ Y-axis is latitude
    - ⊙ Colour is temperature
    - ⊙ 19 different surface temperature images of FK Comae
    - ⊙ Large starspots at high latitudes
    - ⊙ Sometimes two active longitudes
- Light curve mean, amplitude, period and minima are constantly changing.

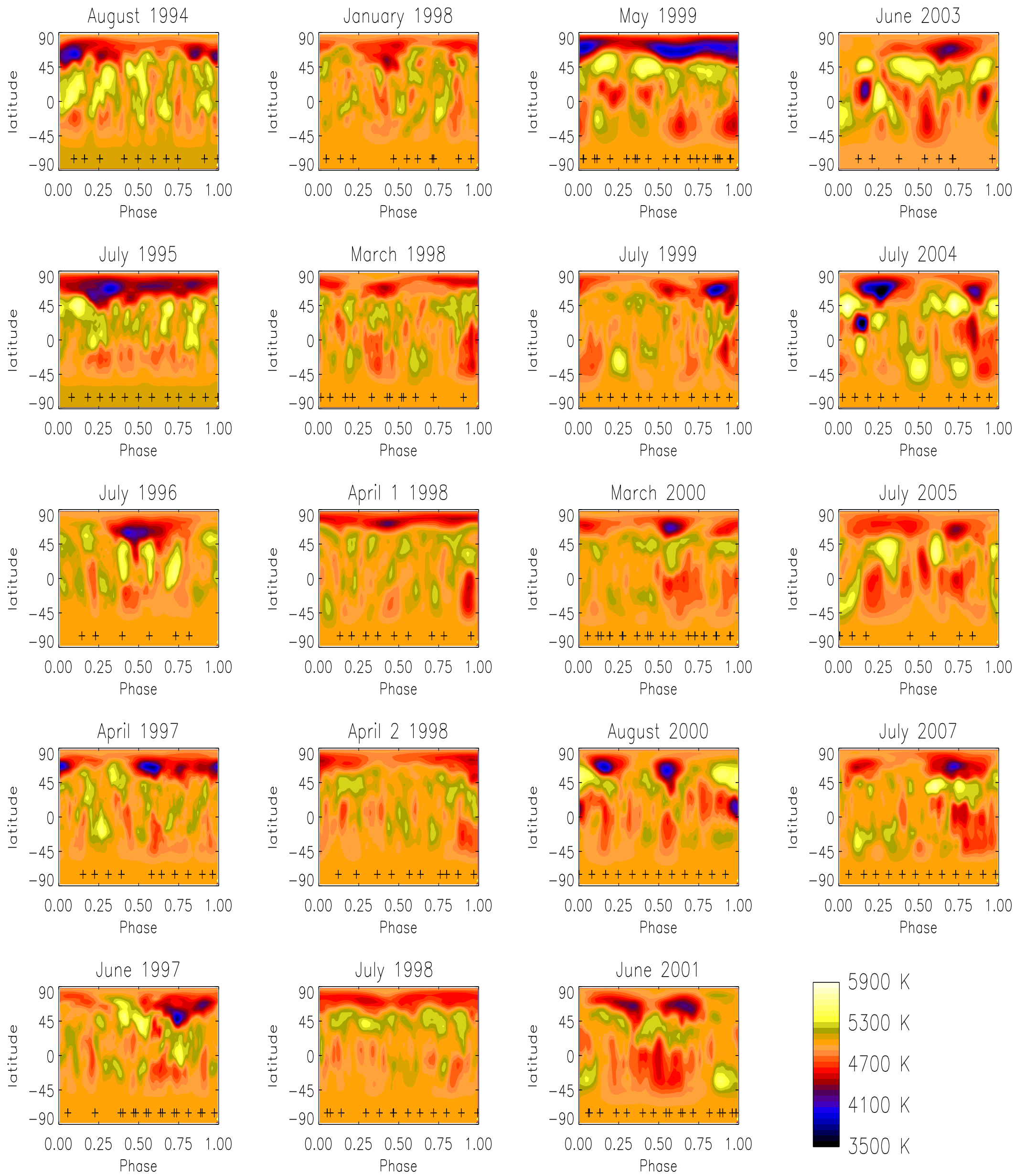

Fig. 2. Korhonen et al. (2009, RMxAC 36, 323): Surface temperature maps of FK Comae

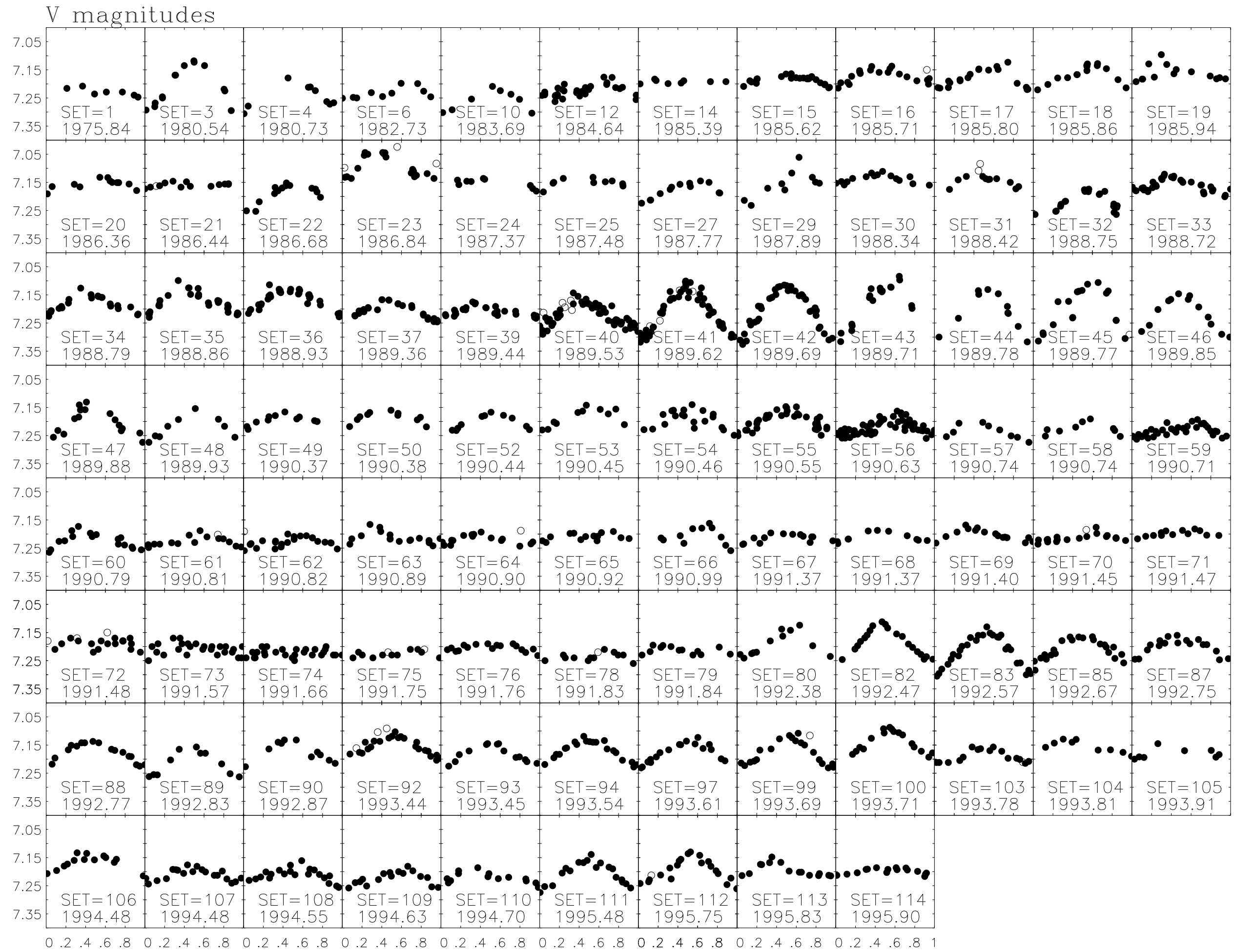

Fig. 3. Jetsu et al. (1999, A&A 351, 212): HD199178 light curves

- Case study: HD199178
  - ⊙ First study in 1990: power spectrum analysis
  - ⊙ Next study in 1999: three stage period analysis
  - ⊙ Rapidly rotating single giant, $P \approx 3.^{\rm d}3$
  - ⊙ Light curve changes within a month
  - ⊙ Separately modelled datasets
  - ⊙ Amplitude from zero to $0.^{\rm m}2$
  - ⊙ Period variations of about 7% (Fig. 4 next page)
  - ⊙ No regular activity cycle (Fig. 5 next page)
  - ⊙ Most light curves are not simple sinusoids
  - ⊙ Some light curves have two minima (e.g. SET=72)

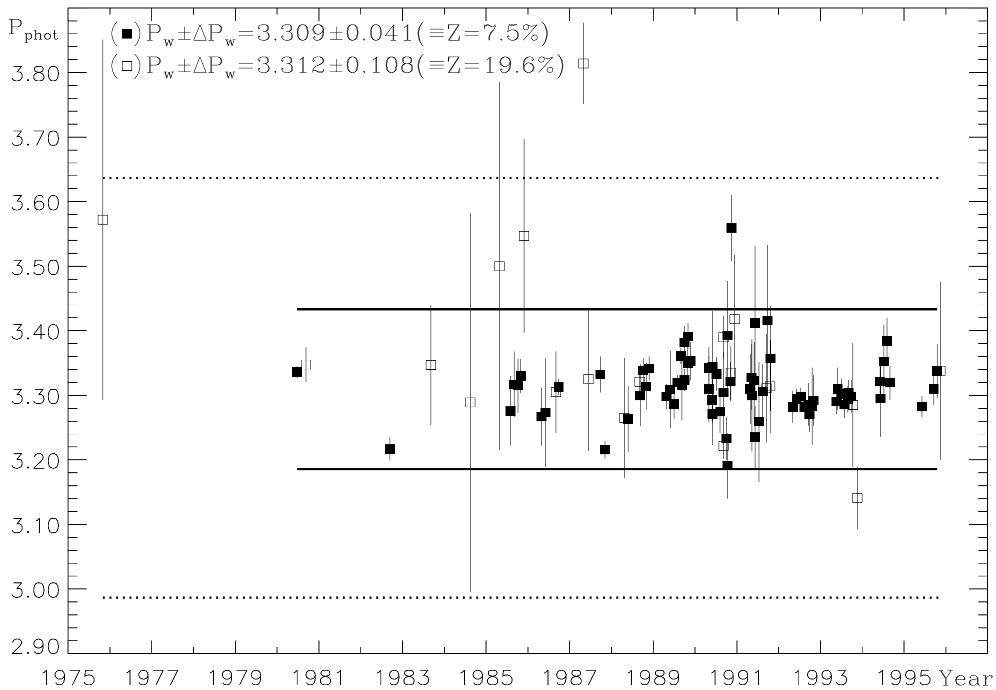

Fig. 4. Light curve periods of HD199178

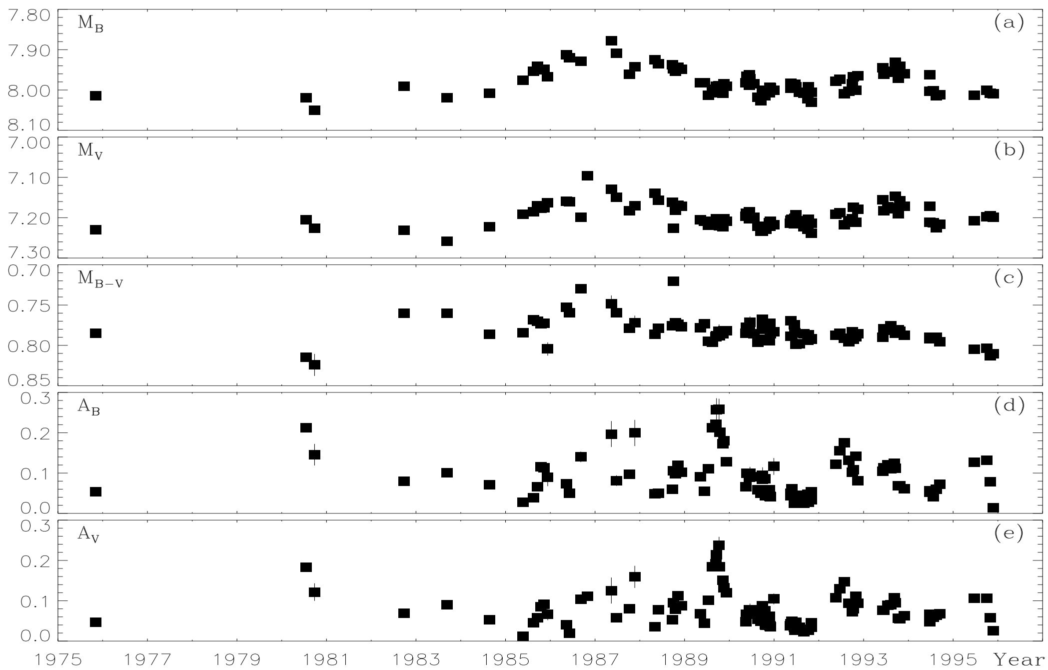

Fig. 5. B and V light curve mean and amplitude of HD199178

3. Unevenly spaced data

- Irregularly spaced observations in time, e.g. astronomy
  - ⊙ Effects caused by irregular gaps can not be removed.
- Evenly spaced observations in time, e.g. economy
  - ⊙ Effects caused by regular gaps can be removed.
- Concepts
  - ⊙ n = Number of observations
  - ⊙ t_{i} = Observing times $(i = 1, ..., n)$
  - ⊙ $y_{\mathrm{i}} = y(t_{\mathrm{i}})$ = Values of observations $(i = 1, ..., n)$
  - ⊙ σ_{i} = Accuracy of observations $(i = 1, ..., n)$
  - ⊙ $w_{\mathrm{i}} = \sigma_{\mathrm{i}}^{-2}$ = Weights of observations $(i = 1, ..., n)$
  - ⊙ t_0 = Zero epoch in time
- Vector notations
  - ⊙ $\bar{t} = [t_1, t_2, ..., t_{\mathrm{n}}]$
  - ⊙ $\bar{y} = [y_1, y_2, ..., y_{\mathrm{n}}]$
  - ⊙ $\bar{\sigma} = [\sigma_1, \sigma_2, ..., \sigma_{\mathrm{n}}]$
  - ⊙ $\bar{w} = [w_1, w_2, ..., w_{\mathrm{n}}]$
- P = Period
- $f = 1/P$ = Frequency
- $\phi_{\mathrm{i}} = \mathrm{FRAC}\,[(t_{\mathrm{i}} - t_0)/P] = \mathrm{FRAC}\,[f(t_{\mathrm{i}} - t_0)]$ = Phase
  - ⊙ FRAC[x] removes integer part of argument $x \Rightarrow 0 \leq \phi_{\mathrm{i}} < 1$
- Problem: Determine the period (P) for the time series $\bar{y} \pm \bar{\sigma}$.
- The introduced methods can be applied to any type unevenly spaced data.
- Examples of real unevenly spaced data will be mostly photometry of variable stars.
- Simulated unevenly spaced data is used in the assignments.

4. Power spectrum method

- Probably the most frequently applied period test in astronomy.
- Revised version in Scargle (1982, ApJ 263, 853: his Eq. 10).
- The power spectrum of the data $y_i = y(t_i)$ for any arbitrary tested frequency f is

$$z_{\mathrm{LS}}(f) = \frac{\{\sum_{\mathrm{i=1}}^{\mathrm{n}} y'_{\mathrm{i}} \cos[2\pi f(t_{\mathrm{i}} - \tau)]\}^2}{2\sum_{\mathrm{i=1}}^{\mathrm{n}}\{\cos[2\pi f(t_{\mathrm{i}} - \tau)]\}^2} + \frac{\{\sum_{\mathrm{i=1}}^{\mathrm{n}} y'_{\mathrm{i}} \sin[2\pi f(t_{\mathrm{i}} - \tau)]\}^2}{2\sum_{\mathrm{i=1}}^{\mathrm{n}}\{\sin[2\pi f(t_{\mathrm{i}} - \tau)]\}^2}, \tag{1}$$

where $\hat{y_i} = [\sum y_i]/n$, $y'_i = y_i - \hat{y_i}$ and τ is defined by

$$\tan(4\pi f\tau) = \left[\sum_{\mathrm{i=1}}^{\mathrm{n}} \sin(4\pi f t_{\mathrm{i}})\right]\left[\sum_{\mathrm{i=1}}^{\mathrm{n}} \cos(4\pi f t_{\mathrm{i}})\right]^{-1}.$$

- This test is most sensitive to periodic sinusoidal variations.
- The limitations of this test will be illustrated $\boxed{\text{LATER}}$.
- Note: Always subtract the mean $\hat{y}_{\mathrm{i}}$ from the data y_{i}, i.e. calculate z_{ls} for $y'_i = y_i - \hat{y}_i$.
- Note: If you do not subtract the mean, you get a result, but most probably the wrong result. Beware of "black boxes".
- The test is usually made within the chosen period interval P_{min} and P_{max}.
- This corresponds to the tested frequency interval $f_{\mathrm{min}} = P_{\mathrm{max}}^{-1}$ and $f_{\mathrm{max}} = P_{\mathrm{min}}^{-1}$.
- Do this test (and all other time series analysis tests) with frequences. The reasons are illustrated $\boxed{\text{LATER}}$.
- A suitable step in tested frequencies is

$$f_{\mathrm{step}} = f_0/OFAC, \tag{2}$$

where $f_0 = [\Delta T]^{-1} = [t_n - t_1]^{-1}$ is the distance between two independent tested frequencies and OFAC=10 (i.e. Over Filling Factor).

- The best period $P_{\mathrm{best}} = f_{\mathrm{best}}^{-1}$ for the data gives the highest value of $z_{LS}(f)$, i.e. the maximum of the periodogram is $\max[z_{LS}(f)] = z_{LS}(f_{\mathrm{best}})$.
- **Exercises 1, 2 and 3** will clarify the power spectrum analysis.

  ⊙ Our timetable is flexible. The aim is to go through these exercises during the first 3x2 hours of lectures.

  ⊙ Learning IDL takes time. This determines the pace.
- You will have many programs after this course. Use

  ⊙ Solution for Exercise 1 is `EXERCISE1.PRO`,

  ⊙ Solution for Exercise 2 is `EXERCISE2.PRO`,

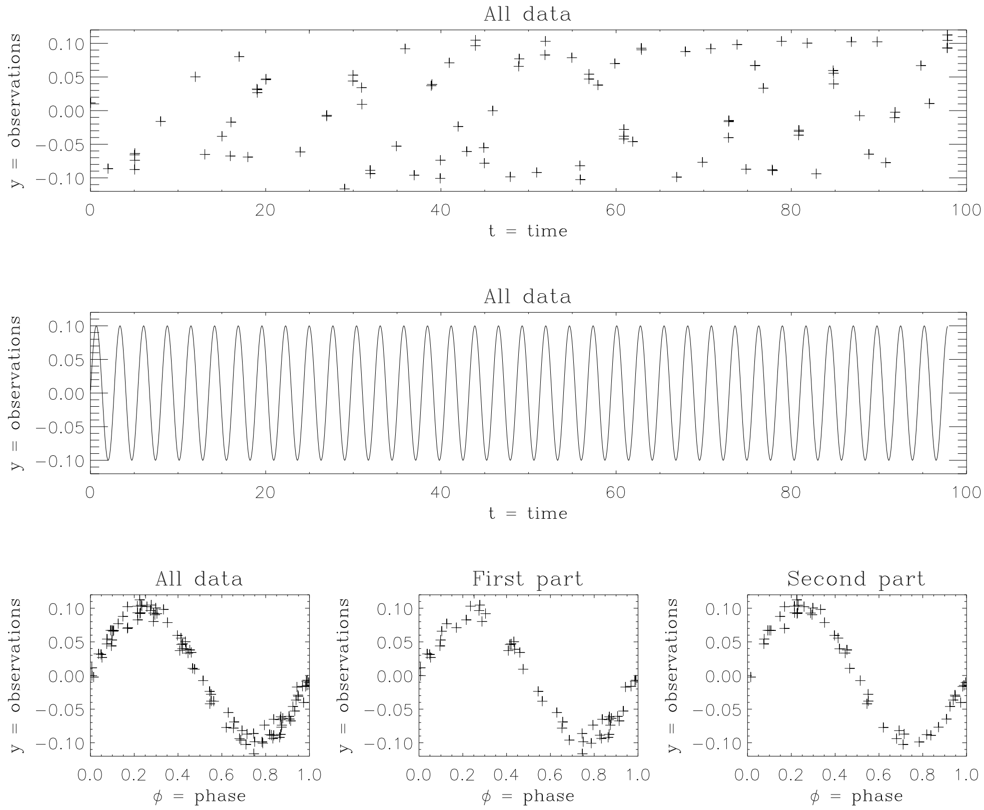

Fig. 6. Data: $\bar{y} = A \sin(2\pi\bar{t}/P) + \bar{\sigma}$, where $A = 0.^{\mathrm{m}}1$, $P = 2.^{\mathrm{d}}7$ and $\bar{\sigma} = N(0, A/10)$ gaussian.

- Sinusoidal data with a constant period and amplitude

  ⊙ Do you see or detect the period $P = 2.7$ from the crosses of the upper plot?

  ⊙ The phase curves of the first and second part of the data are similar.

  ⊙ The same light curve for both parts of the data.

  ⊙ Resembles the orbital velocity curves of binaries.

  ⊙ Simple sinusoid: mean, amplitude, period and phase constant.

  ⊙ This period could be detected even with the power spectrum method, which is sensitive to sinusoidal variations.

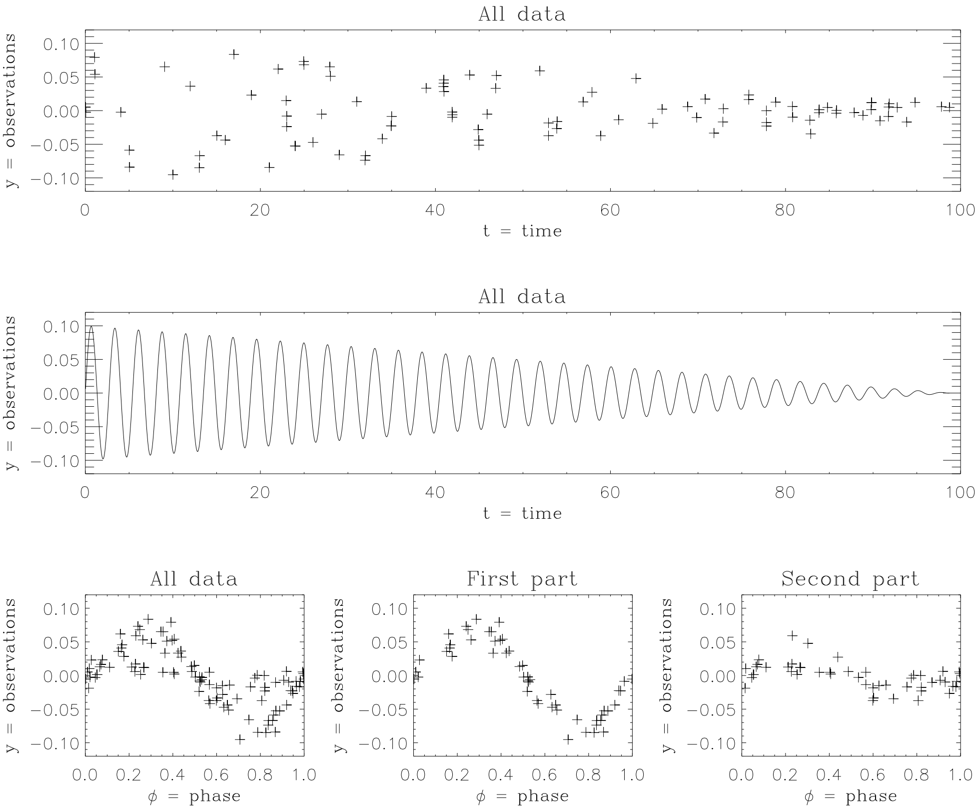

Fig. 7. Data: $\bar{y} = A(t)\sin(2\pi\bar{t}/P) + \bar{\sigma}$, where $A(t) = 0.^{\rm m}1[(t_{\rm n} - t)/(t_{\rm n} - t_1)]$, $P = 2.^{\rm d}7$ and $\bar{\sigma} = N(0, A/10)$ gaussian.

- Sinusoidal data with a constant period, but with decreasing amplitude.

  ⊙ Do you see $P = 2.7$ from the crosses in the upper plot?

  ⊙ The phase curve for all data is messy. The detection of P would be more difficult.

  ⊙ The first and second parts of the data are not similar.

  ⊙ Resembles the case, where a starspot remaining at the same longitude slowly fades away.

  ⊙ Simple sinusoid: mean, period and phase constant. Amplitude changes!

  ⊙ This period might be detected even with the power spectrum method.

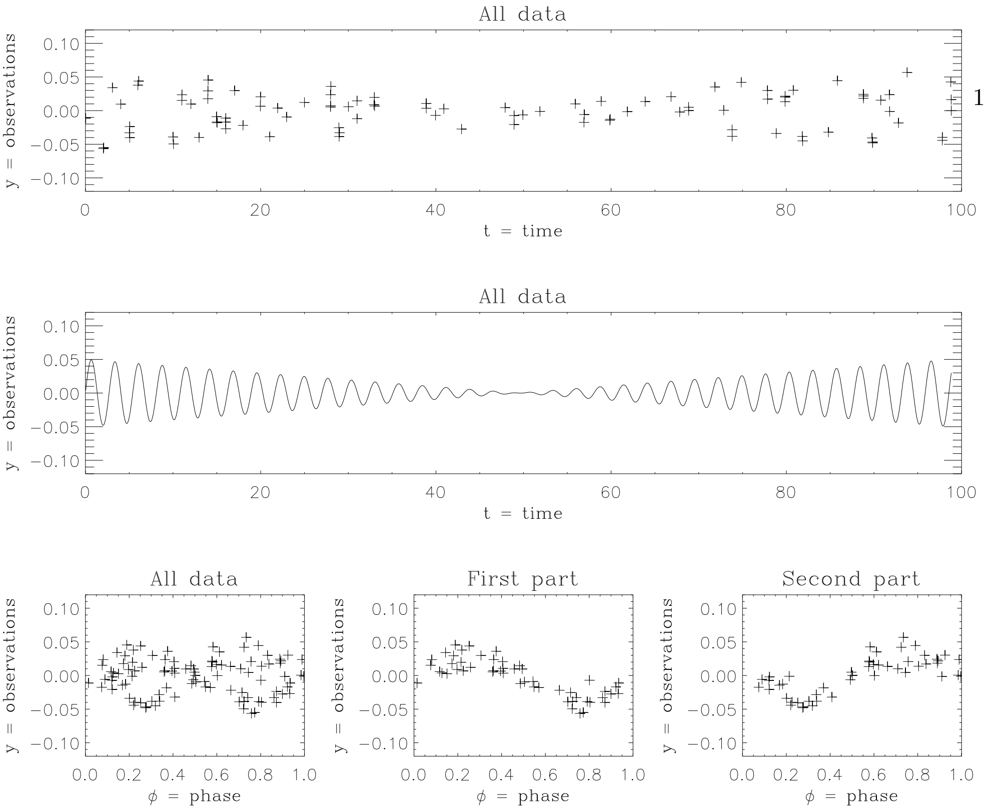

Fig. 8. Data: $\bar{y} = A\sin(2\pi\bar{t}/P) + \bar{\sigma}$, where $A(t) = 0.^{\rm m}1[0.5(t_{\rm n} + t_1) - t)/(t_{\rm n} - t_1)]$, $P = 2.^{\rm d}7$ and $\bar{\sigma} = N(0, A/10)$ gaussian.

- Sinusoidal data with a constant period, but with decreasing and then increasing amplitude.

  ⊙ Do you see $P = 2.7$ from the crosses in the upper plot?

  ⊙ The phase curve for all data is very messy. The detection of P would be very difficult.

  ⊙ The first and second parts of the data are not similar.

  ⊙ A phase shift of 0.5 occurs at $t = 0.5(t_{\rm n} + t_1)$.

  ⊙ Resembles the case, where a starspot remaining at the same longitude slowly fades away. Then another one forms at the opposite side.

  ⊙ Simple sinusoid: mean and period constant.

  ⊙ Amplitude and phase change!

  ⊙ This period would not be detected with the power spectrum method.

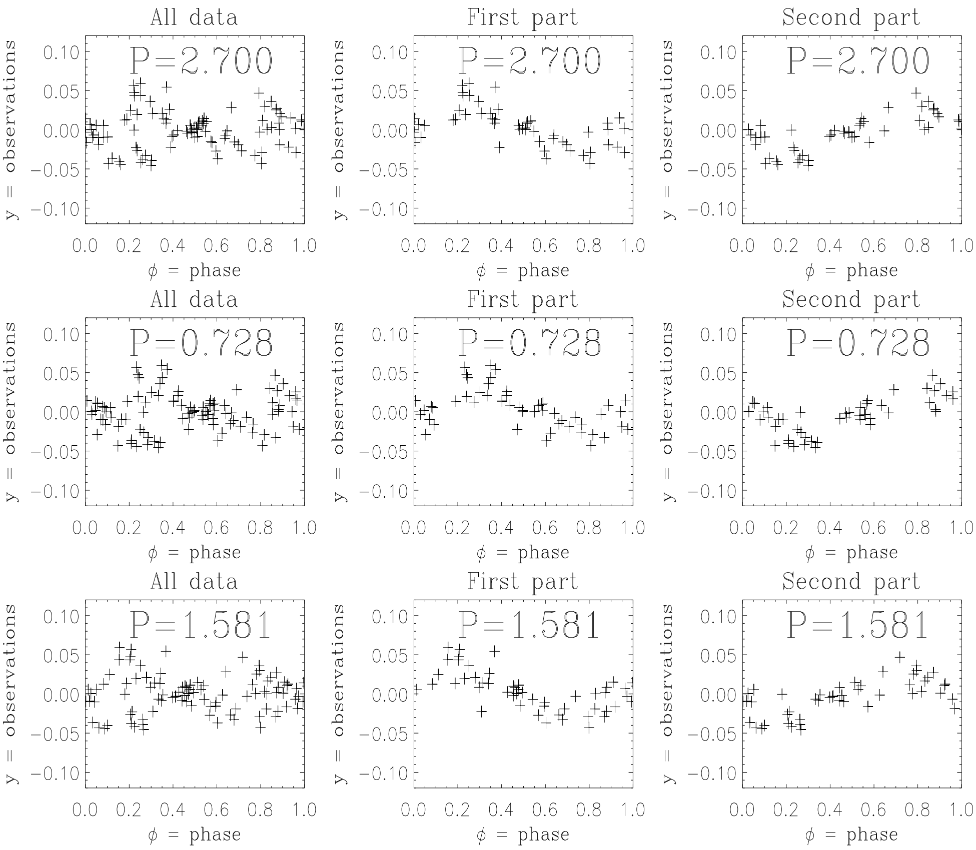

Fig. 9. Data of Fig. 8 with periods $P = 2.^{\rm d}700, 0.^{\rm d}728$ and $1.^{\rm d}581$.

- The same data can be modelled with three different periods. How do we know, what is the correct period?

  ⊙ There are numerous other periods that give nice light curves. The periods $0.^{\rm d}728$ and $1.^{\rm d}581$ were just examples.

  ⊙ If $P = 2.78$ is the real period, then these two other periods are unreal, i.e. spurious periods.

- These data resembled photometry.

  ⊙ Can only be made at night.

  ⊙ $t_{\rm i}$ close to integer multiples of sidereal day $P_0 = 0.^{\rm d}997269$

  ⊙ Simulated $t_{\rm i}$ were $iP_0 \pm 0.^{\rm d}1$ $(i = 1, ..., n)$.

- Sidereal day is one window period of the data.

  ⊙ The time span of data $t_{\rm n} - t_1$ is another window period.

  ⊙ Cloudy nights, seasonal visibility and device failures introduce gaps, i.e. other window periods.

- Problem: A simple sinusoid does not fit the data. The mean, amplitude, period and phase are all changing. The signal to noise ratio A/σ is worse and changing. The window periods introduce unreal, i.e. spurious periods.

5. First encounter with IDL

- `ssh -Y -lusername sky.astro.helsinki.fi` (login)
- `exit` (logout)
- `idl` (begin idl session)
- `IDL>exit` (end idl session)
- Practice the following commands:
  - ⊙ Give the command `idl`
  - ⊙ `IDL>print, dblarr(5)` (Vector with 5 zeros)
  - ⊙ `IDL>print, DblaRR(5)` (Notation not case sensitive)
  - ⊙ `IDL>x=dblarr(5) & y=x+1 & print,y` (3 commands on same line with `&`)
  - ⊙ `IDL>x=[1,2,3,4] & print,x & print,x(0),x(3)` (indices from 0 to n-1)
  - ⊙ `IDL>x=[1,2,3,4] & j=where(x gt 2.5) & print,x(j)` (using indices)
- We shall use `emacs` to edit our first IDL-program.
  - ⊙ Begin editing command: `emacs MODEL0.PRO &`
  - ⊙ Save your current version with the command: select File select Save
  - ⊙ End editing command: ctrl x and ctrl c

– **Write this program MODEL0.PRO. It simply does exactly the same things that were already done during the above IDL-session.**

```
; ________________ Comment line begins with: ; ______________________ ;
; MODEL0.PRO     ; The name of this program, should I forget it.
print, dblarr(5) ; Vector with 5 zeros
print, DblaRR(5) ; Notation not case sensitive
x=dblarr(5) & y=x+1 & print,y   ; 3 commands on same line with &
x=[1,2,3,4] & print,x & print,x(0),x(3)      ; indices from 0 to n-1)
x=[1,2,3,4] & j=where(x gt 2.5) & print,x(j) ; using indices
END ; Every IDL-program ends with "END"
```

- We shall use IDL to run the program `MODEL0.PRO`
  - ⊙ Begin idl with the command: `idl`
  - ⊙ Run the program in idl with the command: `.run MODEL0.PRO`
  - ⊙ End idl with the command `exit`:
- Editing programs and then running them with IDL. We do this with the “trial and error” and “swim or drown” principles.

- Next example: If a file named `MODEL1.DAT` has this two column format and less than 1000 lines,

```
1.022      0.058
1.055      0.090
1.999      0.028
2.052      0.006
.....      .....
```

the data in this file can be read with an IDL-program like this

```
; ________________ Comment line begins with: ; ______________________ ;
; MODEL1.PRO            ; The name of this program, should I forget it.
DATAFILE='MODEL1.DAT'   ; File name is a string marked with: '
X=DBLARR(1000)          & Y=X         & N=0
OPENR,1,DATAFILE        & WHILE NOT EOF(1) DO BEGIN & Q=DBLARR(2)
READF,1,Q & X(N)=Q(0)   & Y(N)=Q(1)   & N=N+1   & ENDWHILE   & CLOSE,1
J=WHERE(X GT 0.)        & X=X(J)      & Y=Y(J)
END                     ; Every IDL program ends with the command END
; ___________________________________________________________________;
```

- Learning a new language is always difficult. But we go through the above example in MODEL1.PRO, until it gets boring.

- In the next example, we want to calculate the sum

$$v(f_j) = \sum_{i=1}^{n} \cos 2\pi f_j t_i, \tag{3}$$

where $t_i = 1, 2, ..., n$, $n = 10$, $f_j = 1, 2, ...m$ and $m = 5$. This program is one version of the solution

```
; ----------------------------------------------------------------- ;
; MODEL2.PRO           ; The name of this program, should I forget it.
; ----------------------------------------------------------------- ;
N=10 & T=1.D0+FINDGEN(10) & PRINT,'I want to check what is t=',t
M=5  & F=1.D0+FINDGEN(5)  & PRINT,'I want to check what is f=',f
V=0.D0*F                            ; v(f) and f have the same length.
FOR A=0,N_ELEMENTS(F)-1 DO BEGIN    ; Loop calculating v(f) begins
V(A)=TOTAL(COS(2.D0*!PI*F(A)*T))    ; Note: !PI=3.14
ENDFOR                              ; Loop -"-             ends
PRINT,'I what to check what is v(f)=',v
PRINT,'I (most probably) could have computed this even without IDL.'
; How to compute phases (Useful in Exercise 1)
TIME=FINDGEN(5)+0.5 & PERIOD=5.6   ; Time, period, phase ...
PHI=TIME/PERIOD     & PRINT,PHI    ; Rounds not removed ...
PPHI=PHI-FIX(PHI)                  ; Rounds removed, FRAC[] example
; How to find maximum of a vector (Useful in Exercise 1)
J=WHERE(TIME EQ MAX(TIME)) & SUURIN=TIME(J) & PRINT,SUURIN
END
; ----------------------------------------------------------------- ;
```

- An example of how to find a maximum (i.e. `max`) and how to compute phases (i.e. `fix`) was also given.
- Again, we go through the above example in MODEL2.PRO, until it gets boring.

- Next example: We want to plot the two functions

  $w(t) = cos(2\pi tf)$, where $f = P^{-1}$, $P = 1.5$, $0 \leq t \leq 5$

  $g(u) = e^{-u}$, where $-2 \leq u \leq 2$

  in **separate** panels with continuous lines and then **together** in the same panel.

- This program does the job

```
; ------------------------------------------------------------------ ;
; MODEL3.PRO         ; The name of this program, should I forget it.
; ------------------------------------------------------------------ ;
!P.NOERASE=1         ; Do not erase screen after each new PLOT command
T=FINDGEN(101)/100.D0           ; t=0.00, 0.01, ...,  1.00
T=5.D0*T                        ; t=0.00, 0.05, ...,  5.00
P=1.5 & F=P^(-1.D0)             ; P fixed, f computed
W=COS(2.D0*!PI*T*F)             ; W(t)
SET_VIEWPORT,0.1,0.4,0.6,0.9    ; Panel location on screen/paper
PLOT,T,W                        ; Continuous line: PSYM=0 is default
OPLOT,T,W,PSYM=1                ; Overplot with crosses: PSYM=1
U=FINDGEN(101)/100.D0           ; u=0.00, 0.01, ...,  1.00
U=4.D0*U-2.D0                   ; u=-2.0,-1.96, ...,1.96,2.0
G=EXP(-1.*U)                    ; g(u)
SET_VIEWPORT,0.6,0.9,0.6,0.9    ; Panel location on screen/paper
PLOT,U,G                        ; Continuous line: PSYM=0 is default
OPLOT,U,G,PSYM=4                ; Overplot with diamonds: PSYM=4
; The X-axis and Y-axis limits are different in the two panels.
X1=MIN([T,U]) & X2=MAX([T,U])   ; Suitable X-axis limits
Y1=MIN([W,G]) & Y2=MAX([W,G])   ; -"-      Y-axis limits
SET_XY,X1,X2,Y1,Y2              ; XY-axis limits fixed
SET_VIEWPORT,0.1,0.9,0.1,0.4    ; Panel location on screen/paper
 PLOT,T,W,PSYM=0, LINESTYLE=0   ; Continuous line
OPLOT,U,G,PSYM=0, LINESTYLE=1   ; Dotted continuous line
END
; ------------------------------------------------------------------ ;
```

- Are there any problems, if you run this program twice during the same IDL-session?

  What is the problem? Try/learn the commands `help` and `.reset_session`

- Again ... until it gets boring.

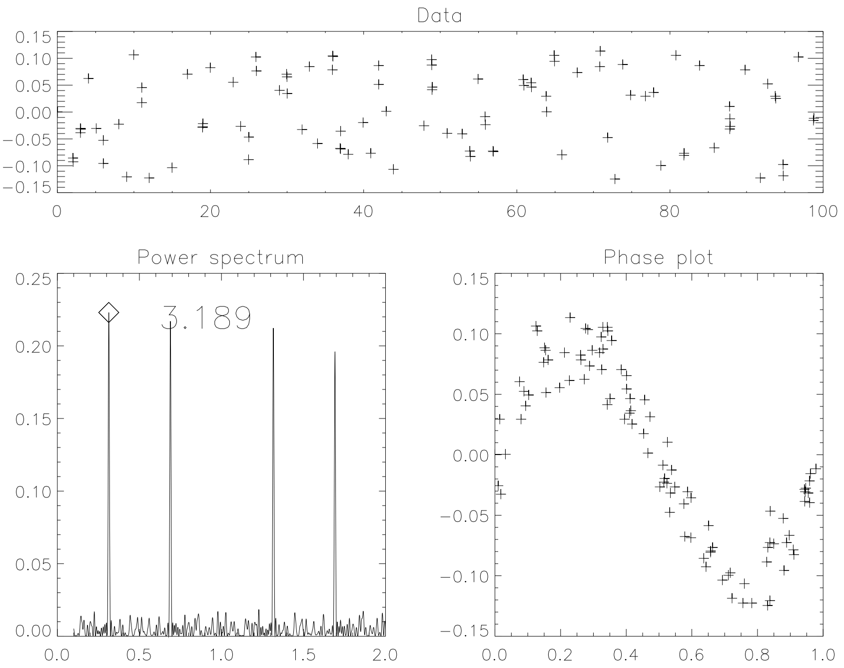

Fig. 10. One example of a solution for **Exercise 1.**

- **Exercise 1.** Write an IDL-program EXERCISE1.PRO calculates the $z_{\rm LS}(f)$ periodogram of Eq. 1 for each of your three datasets `*AADA*.DAT` between $P_{\rm min} = 0.^{\rm d}5$ and $P_{\rm max} = 10^{\rm d}$. Use OFAC = 20 in computing the tested frequencies.

  (a) Plot data y_i as a function of time t_i (1st panel).

  (b) Plot power spectrum $z_{LS}(f_j)$ as a function of tested frequencies f_j (2nd panel).

  (c) Plot data y_i as a function of phase $\phi_i = \mathrm{FRAC}[(t_i - t_1)/P_{\rm best}]$ computed with the best period $P_{\rm best}$ detected with the power spectrum method (3rd panel).

$\sim \bullet \sim$ **Exercise ends here.** $\sim \bullet \sim$

- The overall structure for the program of your **exercise 1**.

  ⊙ MODEL1.PRO example connected to **how** to read your data.

  ⊙ MODEL2.PRO example connected to **how** to calculate $z_{LS}(f)$.

  ⊙ MODEL3.PRO example connected to **how** to plot the results.

6. Linear correlation coefficient

- The linear correlation coefficient between two samples $\bar{x} = [x_1, x_2, ..., x_n]$ and $\bar{y} = [y_1, y_2, ..., y_n]$ is

$$r = r[\bar{x}, \bar{y}] = \frac{\sum_i^n (x_i - m_x)(y_i - m_y)}{(n-1)s_x s_y}, \tag{4}$$

where m_x and s_x are the mean and standard deviation of $\bar{x}$, while m_y and s_y are the mean and standard deviation of $\bar{y}$. The probability for the event that $|r|$ exceeds a fixed value $|r_0| \geq 0$ is

$$P(|r| > |r_0|) = 1 - \mathrm{erf}[|r_0|\sqrt{n/2}], \tag{5}$$

where $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t} dt$ is called the error function. The IDL-command `erf(x)` is used to calculate this error function in MODEL4.PRO below.

- The possible values of the linear correlation coefficient are $-1 \leq r \leq 1$.
  - $\odot$ $r = 1$ means positive linear correlation: all data on an ascending line
  - $\odot$ $r = -1$ means negative linear correlation: all data on a descending line
  - $\odot$ $r = 0$ means no linear correlation: data totally scattered

– This program is a useful model for **exercise 2**:

```
; ------------------------------------------------------------------ ;
; MODEL4.PRO          ; The name of this program, should I forget it.
; ------------------------------------------------------------------ ;
X=[7.,-9.,6.,4.] & Y=[-1.,3.,8.,15]      ; Arbitrary samples X and Y
N=N_ELEMENTS(X)                          ; n
MX=MEAN(X) & SX=STDEV(X)                 ; Mean and standard dev. of X
MY=MEAN(Y) & SY=STDEV(Y)                 ; Mean and standard dev. of Y
R1=TOTAL((X-MX)*(Y-MY))/((N-1.DO)*SX*SY) ; Linear corr. coeff. r
; -The same result could by obtained with the following IDl-command
R2=CORRELATE(X,Y)                        ; Linear coff. coeff. r
PRINT,'R1= ',R1,', R2=',R2               ; Just checking
Q=1.DO-erf(ABS(R1)*SQRT(N/2.DO))         ; Probability P(r>r1)
; -There is also most probably an IDL-routine for this purpose, but...
; ------------------------------------------------------------------ ;
END
; ------------------------------------------------------------------ ;
```

– ... until it gets boring.

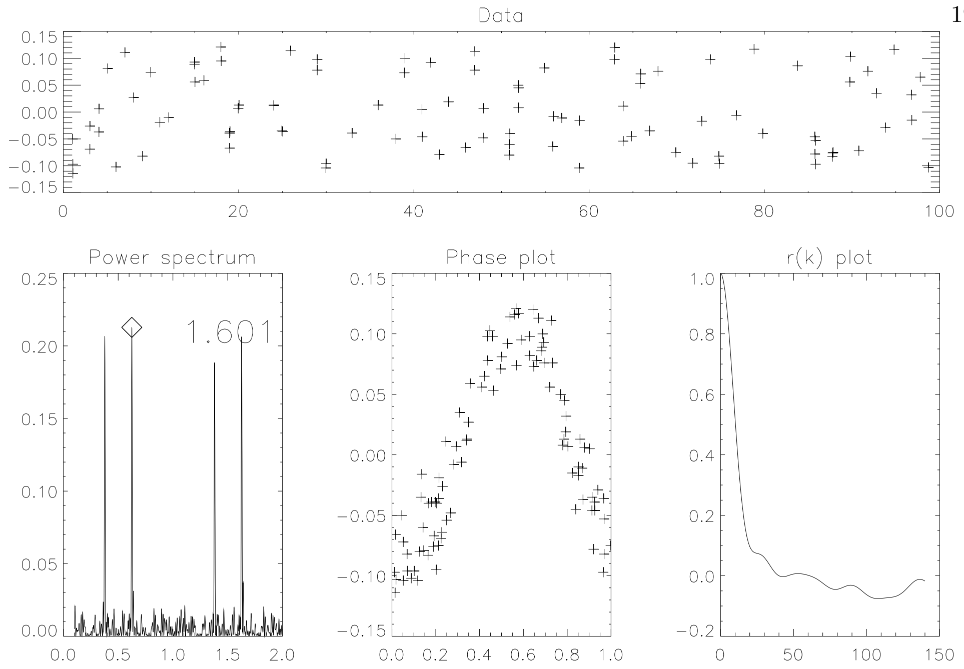

Fig. 11. One example of a solution for **Exercise 2.**

- **Exercise 2.** Choose one of the periodograms $z_{LS}(f_j)$ of your three datasets in Exercise 1. A shorter notation for the M periodogram values is simply $\bar{z} = [z_1, z_2, ..., z_M]$. Choose the following two samples
  $\bar{z_1}(k) = [z_1, z_2, ..., z_{M-k}]$ and $\bar{z_2}(k) = [z_{1+k}, z_{2+k}, ..., z_M]$
  from $\bar{z}$. Note that $\bar{z_1}(k) = \bar{z_2}(k)$ for $k = 0$. Edit an IDL-program EXERCISE2.PRO that computes the linear correlelation coefficient

$$r(k) = r[\bar{z_1}(k), \bar{z_2}(k))] \tag{6}$$

between these two samples for the following values $k = 0, ..., 7$ x OFAC, where OFAC is the overfilling factor value that you used in Exercise 1. Plot $r(k)$ as a function k. The easiest way to solve this exercise is just to add a few lines and one new panel for the $r(k)$ plot into your own former program of Exercise 1.

∼ • ∼ Exercise ends here. ∼ • ∼

- Aims of Exercise 2:

⊙ To understand what linear correlation coefficient is and what it measures.

⊙ To understand that the distance between k and $k+1$ is $f_{\rm step} = f_0/{\rm OFAC}$.

⊙ To understand that the closeby periodogram values are correlated. This result means that the distance between two independent tested frequencies in the power spectrum test with $z_{\rm LS}$ is

$$f_0 = \Delta T^{-1} \Rightarrow (f \pm f_0)\Delta T - f\Delta T = \pm f_0 \Delta T = \pm 1 \quad (7)$$

where $\Delta T = t_n - t_1$ is the time span of the data.

⊙ This means that if f_0 added to <u>any</u> frequency f, the number of full cycles increases with one!

⊙ This means that if f_0 subracted from <u>any</u> frequency f, the number of full cycles decreases with one!

⊙ To understand that the periodogram $z_{LS}(f)$ does not vary much within the frequency interval $\pm f_0/2 = \pm \Delta T^{-1}/2$.

⊙ To understand the consequences of this result. Namely that if a frequency interval between $f_{\rm max}$ and $f_{\rm min}$ is tested, the (integer) number of independent tested frequencies is simply

$$m = {\rm FIX}[(f_{\rm max} - f_{\rm min})/f_0], \quad (8)$$

where FIX$[x]$ removes the decimal part of its argument x.

⊙ To understand why $\Delta T \gg P$ must be filfilled in any reasonable test. The reason is that when ΔT approaches P, the number of independent tested frequencies close to $f = P^{-1}$ approaches one.

⊙ Understand, why is there no sense in testing $P \gtrsim \Delta T$?

⊙ To understand that the power spectrum test (and any other test) must made with frequencies, not with periods. The reason is that if one uses an evenly spaced grid of test periods $P_i = P_{\rm min} + iP_0$, the corresponding grid of tested frequencies $f_i = P_i^{-1} = [P_{\rm min} + iP_0]^{-1}$ is unevenly spaced. This violates the principle of evenly spaced independent tested frequencies, which leads to false statistics.

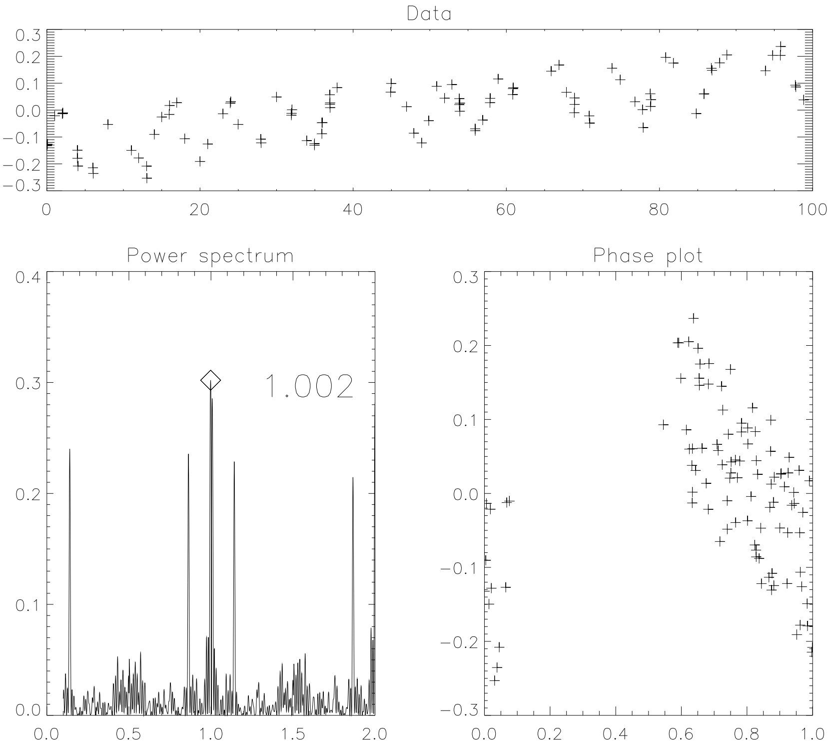

Fig. 12. One example of a solution for **Exercise 3.**

- **Exercise 3.** The power spectrum method finds some periodicity from your three `*AADA*.DAT` datasets $\bar{y}_{\mathrm{old}}$. Test what happens when you add a linear trend to your data. Add the trend $y_{\mathrm{new}}(t_i) = y_{\mathrm{old}}(t_i) + A[(t_i - t_1)/(t_n - t_1)]$, where $A = \max[\bar{y}_{\mathrm{old}}] - \min[\bar{y}_{\mathrm{old}}]$. Compute the power spectrum for $\bar{y}_{\mathrm{new}}$. What happens? Again, the easiest way to solve this exercise is just to add **one line** to your own former program of Exercise 1.
- Aim of Exercise 3: To understand the limitations of the power spectrum method.

7. IDL subroutines

- Subroutines are usefull for several reasons:
  - ⊙ No need to repeat same lines of the program
  - ⊙ No need to edit input/output in the main program
  - ⊙ Less errors in the code
  - ⊙ Same subroutine can be used in other programs
- Subroutines have to be coded before the main program.
- Main program uses only the input and output variables of the subroutine.
- For example, the `MODEL2.PRO` can be revised to subroutine `MYFIRSTSUBROUTINE` that computes the sum

$$v(f_j) = \sum_{i=1}^{n} \cos 2\pi f_j t_i, \tag{9}$$

for any arbitrary combination of $\bar{t} = [t_1, t_2, ..., t_n]$ and $\bar{f} = [f_1, f_2, ..., f_m]$.

```
; ------------------------------------------------------------------ ;
; MODEL5.PRO           ; The name of this program, should I forget it.
; ------------------------------------------------------------------ ;
PRO MYFIRSTSUBROUTINE,T,F,V         ; INPUT: T and F, OUTPUT: V      ;
V=0.D0*F                            ; v(f) and f have the same length;
FOR A=0,N_ELEMENTS(F)-1 DO BEGIN    ; Loop calculating v(f) begins
V(A)=TOTAL(COS(2.D0*!PI*F(A)*T))    ;
ENDFOR & PRINT,V                    ; Loop -"-              ends
ABC='I do not exist in main program'; Not INPUT or OUTPUT !
RETURN & END                        ; Subroutine ends
; ------------------------------------------------------------------ ;
;                                   ; MAIN PROGRAM BEGINS
X=2.D0*!PI*FINDGEN(50)              ; Input T and F names irrelevant!
Y=11.+FINDGEN(10)                   ; Output V name irrelevant!
MYFIRSTSUBROUTINE,X,Y,Z             ; 1st use of subroutine
; ------------------------------------------------------------------ ;
A=COS(Z*7.D0) & B=SIN(X/90.D0)      ;
MYFIRSTSUBROUTINE,A,B,C             ; 2nd use of subroutine
PRINT,ABC                           ; Here the program fails. Why?
END
; ------------------------------------------------------------------ ;
```

- Understand the "fate" of the variable `ABC`.
- "Until it gets boring ... "

8. Pilot Search (PSch)

- The Pilot Search (hereafter PSch) is described in Sect 3.1. of Paper I.
- The following $N = n(n-1)/2$ pairs $(j > i)$ are computed in the PSch

$t_{\mathrm{i,j}} = \mid t_{\mathrm{i}} - t_{\mathrm{j}} \mid$	difference in $\bar{t}$	does not depend on tested f
$y_{\mathrm{i,j}} = [y_{\mathrm{i}} - y_{\mathrm{j}}]^2$	difference in $\bar{y}$	does not depend on tested f
$w_{\mathrm{i,j}} = (w_{\mathrm{i}} w_{\mathrm{j}})(w_{\mathrm{i}} + w_{\mathrm{j}})^{-1}$	weight of $y_{\mathrm{i,j}}$	does not depend on tested f
$\phi_{\mathrm{f,i,j}} = \mathrm{FRAC}[f t_{\mathrm{i,j}}]$	phase difference with tested f	depends on tested f

- **Note:** The pairs $t_{\mathrm{i,j}}$, $y_{\mathrm{i,j}}$ and $w_{\mathrm{i,j}}$ do not depend on f. Therefore, in any sensible program, these pairs are calculated **outside** the loop of tested frequencies.
- **Note:** Only the $\phi_{\mathrm{f,i,j}}$ pairs depend on the tested f. Therefore, only these pairs are calculated **inside** the loop of tested frequencies.
- The pilot search test statistic for any tested frequency f is

$$\Theta_{\mathrm{pilot}}(f) = \left[\sum_{\mathrm{i=1,j=i+1}}^{n-1,n} Z(\phi_{\mathrm{f,i,j}}) W(t_{\mathrm{i,j}}) w_{\mathrm{i,j}} \right]^{-1} \sum_{\mathrm{i=1,j=i+1}}^{n-1,n} Z(\phi_{\mathrm{f,i,j}}) W(t_{\mathrm{i,j}}) w_{\mathrm{i,j}} y_{\mathrm{i,j}}, \tag{10}$$

$$W(t_{\mathrm{i,j}}) = \begin{cases} 1, & D_{\mathrm{min}} < t_{\mathrm{i,j}} < D_{\mathrm{max}} \\ 0, & \text{otherwise} \end{cases}$$

  - Selection of $t_{\mathrm{i,j}}$, i.e. those with “1” selected and those with “0” rejected.

$$Z(\phi_{\mathrm{f,i,j}}) = \begin{cases} 1, & \phi_{\mathrm{f,i,j}} < \tau \\ 1, & \phi_{\mathrm{f,i,j}} > 1 - \tau \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

  - Selection of $\phi_{\mathrm{f,i,j}}$, i.e. those with “1” selected and those with “0” rejected.
- The values of the three adjustable correlation lengths are D_{max}, D_{min} and τ are fixed before the test. Typical values are

$2P_{\mathrm{max}} \lesssim D_{\mathrm{max}} \lesssim 10P_{\mathrm{max}}$	$W(t_{\mathrm{i,j}})$ rejects time differences $t_{\mathrm{i,j}}$ larger than D_{max}.
$D_{\mathrm{min}} \approx 0.9P_{\mathrm{min}}$	$W(t_{\mathrm{i,j}})$ rejects time differences $t_{\mathrm{i,j}}$ smaller than D_{min}.
$0.05 \lesssim \tau \lesssim 0.25$	$Z(\phi_{\mathrm{f,i,j}})$ rejects phase differences $\phi_{\mathrm{f,i,j}}$ larger than τ.

- The pilot search periodogram of Eq. 10 looks complicated.
- Another way to express this periodogram is

$$\Theta_{\mathrm{pilot}}(f) = \left[\sum_{\text{Pairs selected with W or Z}} w_{\mathrm{i,j}} y_{\mathrm{i,j}}\right] / \left[\sum_{\text{Pairs selected with W or Z}} w_{\mathrm{i,j}}\right] = \sum w_{\mathrm{i,j}} y_{\mathrm{i,j}} / \sum w_{i,j}$$

- $\Theta_{\mathrm{pilot}}(f)$ is simply the weighted mean of $y_{i,j}$ selected with $W(t_{\mathrm{i,j}})$ or $Z(\phi_{\mathrm{f,i,j}})$.
- The best period $P_{\mathrm{best}} = f_{\mathrm{best}}^{-1}$ minimizes the squared differences $y_{\mathrm{i,j}}$ that are suitably separated in time and suitably close in phase, i.e. $\min[\Theta(f)] = \Theta(f_{\mathrm{best}})$.
- PSch test is made between P_{min} (f_{max}^{-1}) and P_{max} (f_{min}^{-1})
- The values of $D_{\mathrm{min}}, D_{\mathrm{max}}$ and τ are fixed
- A suitable step in tested frequencies is $\Delta f_{\mathrm{pilot}} = [\mathrm{OFAC}\ D_{\mathrm{max}}]^{-1}$, where a suitable value for the Over Filling Factor is OFAC $= 10$.
- All multiples $f_{\mathrm{j}} = f_{\mathrm{min}} + j\Delta f_{\mathrm{pilot}}$ ($j = 0, 1, 2...$) between f_{min} and f_{max} are tested.
- Imagine any periodic continuous curve. If the observations follow such a curve, the differences between observations close in phase are small, while the differences between observations far in phase are large.

  ⊙ For $K = 1$, draw a sinusoid $y_{\mathrm{i}} = \sin 2\pi K\phi_{\mathrm{i}}$ in the phase interval $0 \leq \phi_{\mathrm{i}} < 1$. Look at your drawing and figure out what would be a suitable value for τ in this case?

  ⊙ For $K = 2$, draw a sinusoid $y_{\mathrm{i}} = \sin 2\pi K\phi_{\mathrm{i}}$ in the phase interval $0 \leq \phi_{\mathrm{i}} < 1$. Look at your drawing and figure out what would be a suitable value for τ in this case?

- Before going to Exercises 4, 5 and 6 ...
- Here is one logical structure for the PSch program.

	Task	Math and/or programming
1	Input	$t_{\rm i}$, $y_{\rm i}$, $\sigma_{\rm i}$, τ, $P_{\rm min}$, $P_{\rm max}$, Q_1, Q_2
2	Output	$\Delta f_{\rm pilot}$, f, $\Theta_{\rm pilot}(f)$, Best periods
3	Derive	$t_{\rm i,j}, y_{\rm i,j}, w_{\rm i,j}, D_{\rm min} = Q_1 P_{\rm min}$, $D_{\rm max} = Q_2 P_{\rm max}$, $\Delta f_{\rm pilot} = [10 D_{\rm max}]^{-1}$, $f_{\rm min} = P_{\rm max}^{-1}$, $f_{\rm max} = P_{\rm min}^{-1}$
4	Apply $W(t_{\rm i,j})$	Select $D_{\rm min} < t_{\rm i,j} < D_{\rm max}$ and the respective $y_{\rm i,j}$ and $w_{\rm i,j}$.
5	Derive all tested f	All integer multiples of $\Delta f_{\rm pilot}$ between $f_{\rm min}$ and $f_{\rm max}$. Create a vector $\Theta_{\rm pilot}$ of same length.
6	Begin f loop	
7	Apply $Z(\phi_{\rm f,i,j})$	Select all $\phi_{\rm i,j,f} < \tau$ or $\phi_{\rm i,j,f} > 1 - \tau$ and the respective $y_{\rm i,j}$ and $w_{\rm i,j}$
8	Compute $\Theta_{\rm pilot}(f)$	$\Theta_{\rm pilot}(f) = \frac{\sum_{\rm i=1,j=i+1}^{n-1,n} w_{\rm i,j} y_{\rm i,j},}{\sum_{\rm i=1,j=i+1}^{n-1,n} w_{\rm i,j}}$ for the selected $y_{\rm i,j}$ and $w_{\rm i,j}$.
9	End f loop	
10	Interpret $\Theta_{\rm pilot}(f)$	Graphical presentation and identification of best P.

- “Until it gets boring ...”

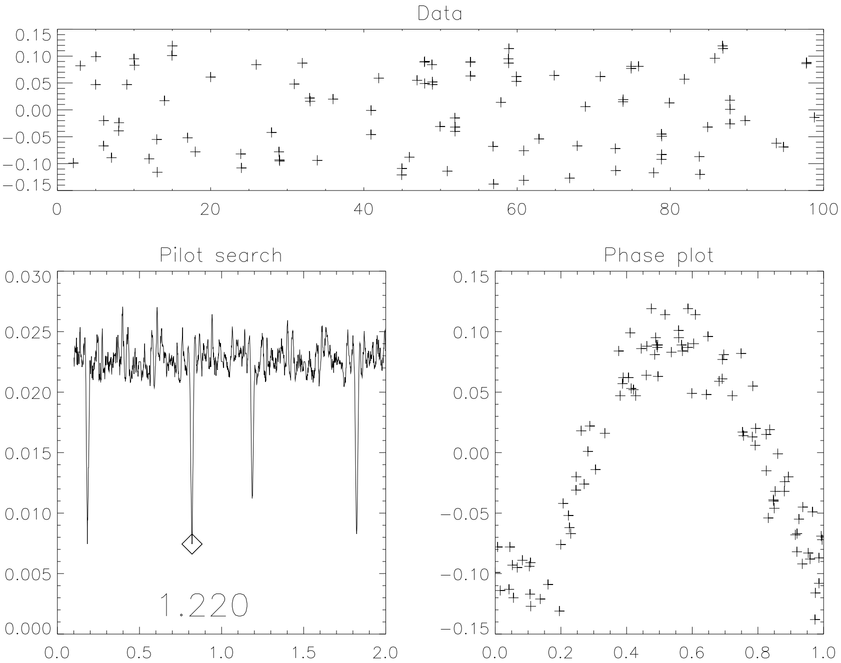

Fig. 13. One example of a solution for **Exercise 4.**

- **Exercise 4.** Write an IDL-program EXERCISE4.PRO that calculates the $\Theta_{\rm pilot}(f)$ periodogram of Eq. 10 for each of your three datasets `*AADA*.DAT` between $P_{\rm min} = 0.^{\rm d}5$ and $P_{\rm max} = 10^{\rm d}$. Fix the other parameters of your test to $\tau = 0.25$, $\mathrm{OFAC} = 20$, $D_{\rm min} = 0.9P_{\rm min}$ and $D_{\rm max} = 4P_{\rm max}$. Solve w_i by using $\sigma_i = 0.01$ for errors.

  (a) Plot data y_i as a function of time t_i (1st panel).

  (b) Plot periodogram $\Theta(f_j)$ as a function of tested frequencies f_j. Mark the deepest minimum, e.g. with a diamond (2nd panel).

  (c) Plot data y_i as a function of phase $\phi_i = \mathrm{FRAC}[(t_i - t_1)/P_{\rm best}]$ computed with the best period $P_{\rm best}$ detected with your PSch test (3rd panel).

$\sim \bullet \sim$ **Exercise ends here.** $\sim \bullet \sim$

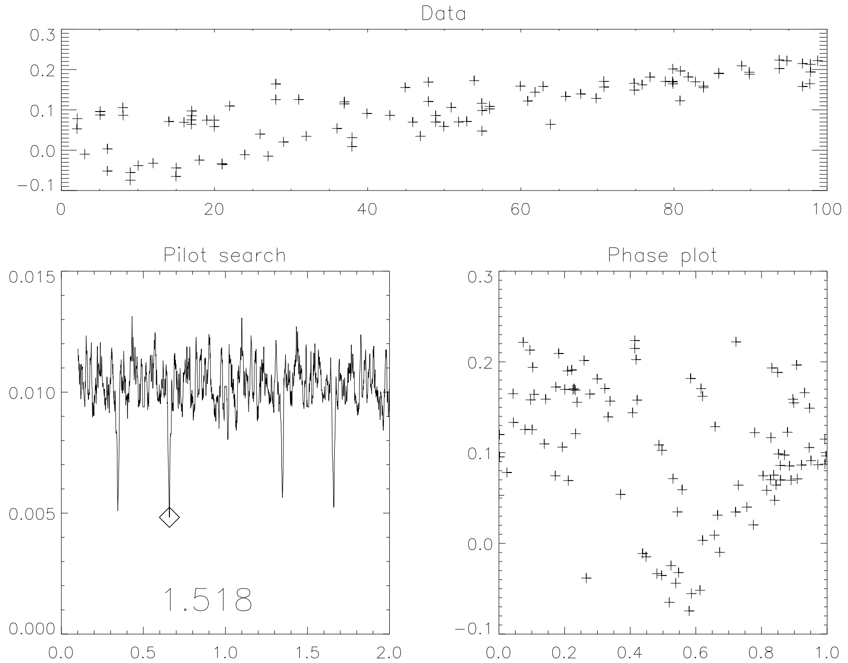

Fig. 14. One example of a solution for **Exercise 5.**

- **Exercise 5.** Test what happens when you add a linear trend to your data $\bar{y}_{\rm old}$ in each of your three datasets `*AADA*.DAT`. Add the trend $y_{\rm new}(t_i) = y_{\rm old}(t_i) + A[(t_i - t_1)/(t_n - t_1)]$, where $A = \max[\bar{y}_{\rm old}] - \min[\bar{y}_{\rm old}]$. Then do the same PSch tests as in exercise 4. Note that you can solve this exercise by adding only one or two lines to the program that you used in Exercise 4.

$\sim \bullet \sim$ **Exercise ends here.** $\sim \bullet \sim$

- Aim of Exercise 5: What do we learn by comparing the results of Exercises 3 and 5?

- **Exercise 6.** The same data were analysed in Exercises 1, 2, 3, 4 and 5. You now get new data `*AAMU*.DAT`. Can you find periodicity between $P_{min} = 0.5$ and $P_{max} = 10$ in these data with the PSch? Note that these data files have three columns: $\bar{t}$=time, $\bar{y}$=observations and $\bar{\sigma}$=error. You therefore have to modify the part of your solution EXERCISE6.PRO so that it reads the data and also uses the weights $w_{\rm i} = \sigma_{\rm i}^{-2}$.

$\sim \bullet \sim$ **Exercise ends here.** $\sim \bullet \sim$

- Unfortunately, an example of a solution would make this Exercise 6 useless.
- Aim of Exercise 6: To understand the flexibility of PSch.

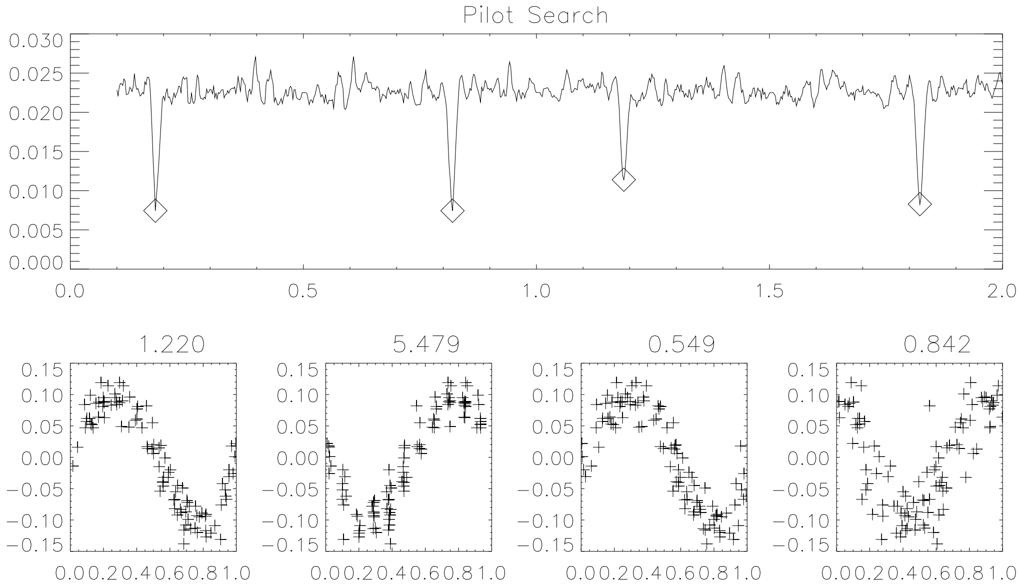

Fig. 15. One example of a solution for **Exercise 7.**

- **Exercise 7.** Edit an IDL-program EXERCISE7.PRO that uses a PSch subroutine: `PILOTSEARCH,T,Y,W,PMIN,PMAX,OFAC,TAU,F,PTHETA,F4,PTHETA4,DMAX`, where

INPUT		OUTPUT	
`T`	Time $\bar{t}$	`F`	Tested $\bar{f}$
`Y`	Observations $\bar{y}$	`PTHETA`	Periodogram $\Theta(f)$
`W`	Weights $\bar{w}$	`F4`	Four best frequencies f_4
`PMIN`	Minimum tested $P_{\min}$	`PTHETA4`	Periodogram values $\Theta(f_4)$
`PMAX`	Maximum tested $P_{\max}$	`DMAX`	D_{max}
`OFAC`	Overfilling factor OFAC		
`TAU`	τ in Eq. 11		

Use the same data, and $D_{\min}$ and $D_{\max}$ values, as in Exercise 4. This program should identify the four best frequencies/periods and then plot the data as function of phase derived with these frequencies/periods using four separate panels. Mark the detected values on the periodogram and give the detected periods above the phase plot (see the model solution). The easiest solution is to copy your `EXERCISE4.PRO` to `EXERCISE7.PRO` and then edit the necessary changes. Test this program to three datasets of Exercise 1. Try different $P_{\min}$, $P_{\max}$, τ and `OFAC` values.

$\sim \bullet \sim$ **Exercise ends here.** $\sim \bullet \sim$

- Aim: To be able to combine later all subroutines of TSPA into one flexible program.

9. Three Stage Period Analysis (TSPA) model

- The TSPA model for the data $\bar{y} \pm \bar{\sigma}$ is

$$g(\bar{\beta}) = g(t, \bar{\beta}) = M + \sum_{\mathrm{k}=1}^{\mathrm{K}} B_{\mathrm{k}} \cos(k2\pi f t) + C_{\mathrm{k}} \sin(k2\pi f t),$$

where $\bar{\beta} = [M, B_1, ..., B_{\mathrm{K}}, C_1, ..., C_{\mathrm{K}}, f]$ denotes the free parameters of the model.

- These $Q = 2K + 2$ free parameters are:
  - ⊙ M = mean
  - ⊙ B_{k} and C_{k} = amplitudes
  - ⊙ f = frequency
- Problem: Find the best values of $\bar{\beta}$ that minimize the residuals $\epsilon_{\mathrm{i}} = y_{\mathrm{i}} - g(t_i, \bar{\beta})$
  - ⊙ Small residuals ϵ_{i} ≡ Reasonable model
  - ⊙ Large residuals ϵ_{i} ≡ Unreasonable model
- The test statistic is

$$\chi^2(\bar{\beta}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} w_{\mathrm{i}} \epsilon_{\mathrm{i}}^2 = \sum_{\mathrm{i}=1}^{\mathrm{n}} \left[\frac{\epsilon_i}{\sigma_i}\right]^2 = \sum_{\mathrm{i}=1}^{\mathrm{n}} \left[\frac{y_{\mathrm{i}} - g(t_i, \bar{\beta})}{\sigma_{\mathrm{i}}}\right]^2.$$

  - ⊙ The weights $w_{\mathrm{i}} = \sigma_i^{-2}$ give more "weight" to the more accurate data.
  - ⊙ The errors σ_{i} (accuracy) are always positive.
  - ⊙ The residuals ϵ_{i} can be both negative and positive
- **Question**: Why must any reasonable model have negative and positive ϵ_{i}?
- **Question**: Why must any reasonable model have $\sigma_1 \approx |\epsilon_1|, ..., \sigma_{\mathrm{n}} \approx |\epsilon_{\mathrm{n}}| \Rightarrow \chi^2 \approx n$?
- The number of free parameters Q must fulfill $Q \leq n - 1$. For example, never model two data points $y_1 = y(t_1)$ and $y_2 = y(t_2)$ with $g(t, \bar{\beta}) = A + Bt$, where $\bar{\beta} = [A, B]$.
- If n is number of observations and Q is the number of free parameters, then the parameter $\nu = n - Q$ is called the degree of freedom.
- Probability $P(\chi^2(\nu) < \chi_0)$ can be computed for any combination of n, Q and χ_0^2.

10. Least Squares Fit (LSF)

- LSF theory is not discussed in detail
- LSF has been performed about $10^{n\to\infty}$ times, i.e. it should not pose surprises.
- LSF solution with IDL is thoroughly discussed: you learn CURVEFIT subroutine.
- LSF test statistic:

$$\chi^2(\bar{\beta}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} w_{\mathrm{i}}\epsilon_{\mathrm{i}}^2 = \sum_{\mathrm{i}=1}^{\mathrm{n}} \left[\frac{y_{\mathrm{i}} - g(t_i,\bar{\beta})}{\sigma_{\mathrm{i}}}\right]^2$$

- LSF solves $\bar{\beta}$ by minimizing χ^2
- LSF minimizes the weighted sum of the residuals, i.e. the difference between model and data.
- LSF with a <u>linear</u> model has a unique $\bar{\beta}$ solution.
- LSF with a <u>nonlinear</u> model does not have a unique $\bar{\beta}$ solution. The $\bar{\beta}$ result depends on the LSF trial solution $\bar{\beta}_{\mathrm{trial}}$.
- <u>Definition:</u> A model $\bar{g}$ is linear, if all partial derivatives $\partial g/\partial\beta_{\mathrm{i}}$ are independent of all β_{j} $(1 \leq i,j \leq Q)$.
- Examples of linear and nonlinear models:

⊙ Linear model: $g = Ax + B$, $\bar{\beta} = [A,B]$, $\partial g/\partial A = x$, $\partial g/\partial B = 1$

⊙ Linear model: $g = M + A\sin x + B\cos x$, $\bar{\beta} = [M,A,B]$, $\partial g/\partial M = 1$, $\partial g/\partial A = \sin x$, $\partial g/\partial B = \cos x$

⊙ Nonlinear model: $g = A^2x + B$, $\bar{\beta} = [A,B]$, $\partial g/\partial A = 2Ax$, $\partial g/\partial B = 1$

Linearization: substitute $A^2 = C$, and model $g = Cx + B$

⊙ Nonlinear model: $g = A\mathrm{e}^{\mathrm{Bx}}$, $\bar{\beta} = [A,B]$, $\partial g/\partial A = \mathrm{e}^{\mathrm{Bx}}$, $\partial g/\partial B = AB\mathrm{e}^{\mathrm{Bx}}$

Linearization: $y' = \ln(y)$, $g' = \ln(g) = \ln A + Bx = C + Bx$, where $C = \ln(\mathrm{A})$

⊙ Nonlinear model: $g = M + A\sin(x-B)$, $\bar{\beta} = [M,A,B]$, $\partial g/\partial M = 1$, $\partial g/\partial A = \sin(x-B)$, $\partial g/\partial B = A\cos(x-B)$

Linearization: $g = M + A\sin(x-B) = M + C\cos x + D\sin x$

⊙ This can not be linearized: $g = M + A\cos(x/P) + B\sin(x/P)$, $\bar{\beta} = [M,A,B,P]$

11. χ^2 statistics

- Assumption: LSF solved the free parameters $\bar{\beta}$ of the model and gave $\chi_0^2 = \chi^2(\bar{\beta})$.
- Problem: How good is the model $g(\bar{\beta}, \bar{t})$?
- Analytical solution:
  - ⊙ Calculate the probability $P(\chi^2 \leq \chi_0^2)$ for $\nu = n - Q$ degrees of freedom
    - ○ Incomplete Gamma function: $P(a,x) = 1/\Gamma(a) \int_0^{\mathrm{x}} t^{\mathrm{a}-1} e^{-\mathrm{t}} dt \quad (a > 0)$,
    - ○ Gamma function: $\Gamma(a) = \int_0^\infty t^{\mathrm{a}-1} \mathrm{e}^{-\mathrm{t}} dt$,
    - ○ Error function: $\mathrm{erf(x)} = 2/\sqrt{\pi} \int_0^{\mathrm{x}} \mathrm{e}^{-\mathrm{t}^2} \mathrm{dt}$.
    - ○ The connection: $\mathrm{erf}(x) = P(\frac{1}{2}, x^2) \quad (x \geq 0)$
  - ⊙ Finally: $P(\chi^2 \leq \chi_0^2) = P(\frac{\nu}{2}, \frac{\chi_0^2}{2})$
  - ⊙ Conclusion: Program these and we might as well ...
- IDL–solution:
  - ⊙ One IDL command solves all above functions.
  - ⊙ Calculate $\chi_0^2 = \chi^2(\bar{\beta})$ and the degrees of freedom $\nu = n - Q$
  - ⊙ Use command `C=CHISQR_PDF(A,B)`
  - ⊙ Command input: `A` is χ_0^2. `B` is ν.
  - ⊙ Command output: `C` is $P(\chi^2 \leq \chi_0^2) = P(\frac{\nu}{2}, \frac{\chi_0^2}{2})$.
- Reject the model $g(\bar{\beta}, \bar{t})$, if $P(\chi^2 > \chi_0^2) = 1 - P(\chi^2 \leq \chi_0^2) \leq \gamma$, where γ is called the preassigned significance level. Typical values are $\gamma = 0.1, 0.05, 0.01, 0.005$ and 0.001.
- Limitations for χ^2 statistics:
  - ⊙ The error estimates $\bar{\sigma}$ must be correct and have a gaussian distribution
  - ⊙ Avoid "overmodelling", i.e. $Q \to n$.

12. CURVEFIT

- IDL performs LSF with the `CURVEFIT` subroutine
- `CURVEFIT` uses the subroutine `FUNCT`
- `FUNCT` contains the model `F` and its partial derivatives `PDER`
- `FUNCT` can replaced by `FUNCTION_NAME`
- `FUNCTION_NAME` allows several LSF models within one program.
- For example, `MODEL6.PRO` first creates artificial data simulated with the model $g(t_{\rm i}) = M + B_1 \sin f2\pi t_{\rm i} + C_1 \cos f2\pi t_{\rm i}$,
  where $f = 1$ (i.e. $P = f^{-1} = 1$), $\bar{\beta} = [M, B_1, C_1]$, $[0 \leq t_{\rm i} < 4\pi P]$ and $(g_{\rm max} - g_{\rm min})/Z$ determines accuracy σ_i.
- Question: Is this a linear model?
- Command `YFIT=CURVEFIT(X,Y,W,B,EB,FUNCTION_NAME='IDEA')`

`YFIT`	$\bar{g}(t, \bar{\beta}_{\rm final})$	OUTPUT
`X`	$\bar{t}$	INPUT
`Y`	$\bar{y}$	INPUT
`W`	$\bar{w}$	INPUT
`B`	$\bar{\beta}_{\rm trial} \Rightarrow \bar{\beta}_{\rm final}$	INPUT $\Rightarrow$ OUTPUT
`EB`	$\bar{\sigma}_\beta$	OUTPUT
`PDER(*,0)=1.D0`	$\frac{\partial g(t,\bar{\beta})}{\partial M}$	INPUT
`PDER(*,1)=SIN(X)`	$\frac{\partial g(t,\bar{\beta})}{\partial B_1}$	INPUT
`PDER(*,2)=COS(X)`	$\frac{\partial g(t,\bar{\beta})}{\partial C_1}$	INPUT

- Find CURVEFIT with "?" command in IDL.
- Question: Does `CURVEFIT` work with nonlinear models.
- Question: Does `CURVEFIT` also work with linear models?
- Answer/Aim: Understand the role of $\bar{\beta}_{\rm trial}$ in nonlinear and linear models!
- We shall edit `MODEL6.PRO` and test it with different combinations of n and Z

```
; ----------------------------------------------------------------
; MODEL6.PRO  ; The name of this program should I forget it
; ----------------------------------------------------------------
PRO IDEA,X,A,F,PDER                           ; SUBROUTINE
F=0.+A(0)+A(1)*COS(X)+A(2)*SIN(X)             ; g (model)
PDER=DBLARR(N_ELEMENTS(X),3)                  ; Partial derivatives
PDER(*,0)=1.D0                                ; dg/dM
PDER(*,1)=COS(X)                              ; dg/dB_1
PDER(*,2)=SIN(X)                              ; dg/dC_1
RETURN & END                                  ; Ends any subroutine.
; ----------------------------------------------------------------
;                                             ; MAIN PROGRAM
N =4.D0  & READ,'Choose ..... n = ',N         ; n simulated pts
Z=1.D0   & READ,'Choose A/sigma = ',Z         ; S/N=Signal to noise
M =0.D0  & READ,'Choose ..... M = ',M         ; Simulated M
B_1=0.D0 & READ,'Choose ... B_1 = ',B_1       ; Simulated B_1
C_1=0.D0 & READ,'Choose ... C_1 = ',C_1       ; Simulated C_1
X=4.D0*!PI*RANDOMU(S,N)                       ; Simulated X
Y=M+B_1*COS(X)+C_1*SIN(X)                     ; Simulation model
A =MAX(Y)-MIN(Y) & E =(A/Z)                   ; A and sigma range
E =E*RANDOMN(S,N)                             ; Random error
MOVE=RANDOMN(S,N)                             ; Simulated y shift
Y =Y+(1.+MOVE)*E                              ; Simulated data
J=SORT(X) & X=X(J) & Y=Y(J) & E=E(J)          ; Rank order.
; ----------------------------------------------------------------
ERASE & !P.NOERASE=1                          ; PLOT
X1=0           & X2=MAX(X)                    ; X-limits
Y1=MIN(Y-3.*E) & Y2=MAX(Y+3.*E)               ; Y-limits
SET_VIEWPORT,.1,.9,.1,.9                      ; Viewport place
SET_XY,X1,X2,Y1,Y2                            ; XY-limits
PLOTERR,X,Y,E,PSYM=4                          ; Simulated data
U=FINDGEN(101)/100.D0                         ; Simulated model
U=MIN(X)+(MAX(X)-MIN(X))*U                    ; -"-
V=M+B_1*COS(U)+C_1*SIN(U)                     ; -"-
OPLOT,U,V,PSYM=0,LINESTYLE=0                  ; -"-
W=E^(-2.D0) & B=DBLARR(3)                     ; w_i and BETA_0
YFIT=CURVEFIT(X,Y,W,B,EB,FUNCTION_NAME='IDEA') ; LSF
XX=MIN(X)+(MAX(X)-MIN(X))*FINDGEN(101)/100.D0  ;Argument
YY=B(0)+B(1)*COS(XX)+B(2)*SIN(XX)             ; Fitted model
OPLOT,XX,YY,PSYM=0,LINESTYLE=1                ; -"-
; ----------------------------------------------------------------
PRINT,'LSF gives  M   = ',B(0),' +/- ',EB(0); PRINT
PRINT,'LSF gives  B_1 = ',B(1),' +/- ',EB(1); -"-
PRINT,'LSF gives  C_1 = ',B(2),' +/- ',EB(2); -"-
CHI0=TOTAL(((YFIT-Y)/E)^2.D0) & NU=N-3        ; Chi^2, nu=n-Q
NYT=1.D0-CHISQR_PDF(CHI0,NU)                  ; P(CHI^2 > CHI0)
PRINT,'CHI^2 = ',CHI0,' Probability = ',NYT ;
; ----------------------------------------------------------------
; -The nexty line converts CHI0 to string and shows the result.
XYOUTS,X1+0.1*(X2-X1),Y1+0.9*(Y2-Y1),STRING(CHI0,'(F8.1)'),SIZE=2
; ----------------------------------------------------------------
END
; ----------------------------------------------------------------
```

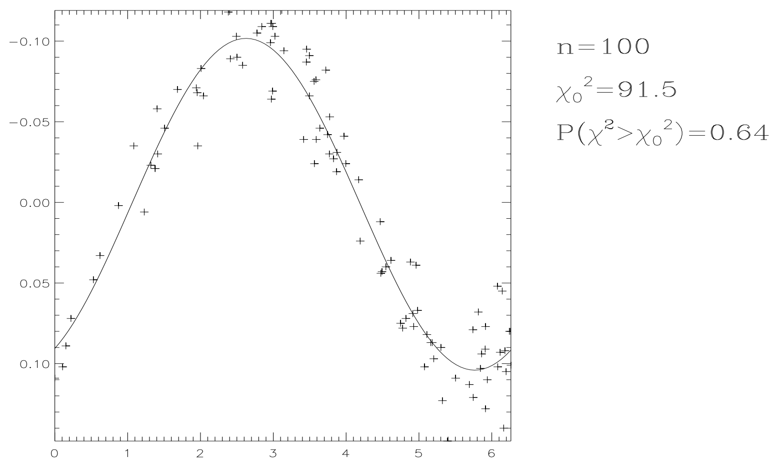

Fig. 16. One model of a solution for Exercise 8.

- **Exercise 8.** Edit an IDL-program EXERCISE8.PRO that solves the following problems: Calculate the n phases ϕ_i with the best period P that you detected in the data y_i of Exercise 4. Calculate $x_i = 2\pi\phi_i$. Fit the data x_i and y_i to the linear model $g(t, \bar{\beta}) = A + B\cos x + C\sin x$, where $x = 2\pi\phi$ and $\bar{\beta} = [A, B, C]$. Solve this model with `CURVEFIT`. Overplot the data and the model. Calculate χ_0^2 with $\sigma_i = 0.02$. Use `XYOUTS` to display the values of n, χ_0^2 and $P(\chi^2 > \chi_0^2)$. The program `MODEL7.PRO` provides an example of the required string manipulations. One model of a solution is shown in Fig. 16.

$\sim \bullet \sim$ **Exercise ends here.** $\sim \bullet \sim$

- For the clumsy fonts `!4` ... of IDL, you can do a web search for the words `"IDL fonts"`.

- **Exercise 9.** Edit an IDL-program EXERCISE9.PRO that solves the following problems: Calculate the n phases ϕ_i with the best period P that you detected in the data y_i of Exercise 6. Note that you have to read the errors σ_i from the data file and use them in the fit and in the calculation of the χ_0^2. Calculate $x_i = 2\pi\phi_i$. Fit the data x_i and y_i to the linear 6th order model model

$$\begin{aligned} g(t, \bar{\beta}) = & A + B\cos x + C\sin x \\ & +D\cos 2x + E\sin 2x \\ & +F\cos 3x + G\sin 3x \\ & +H\cos 4x + I\sin 4x \\ & +J\cos 5x + K\sin 5x \\ & +L\cos 6x + M\sin 6x \end{aligned} \tag{12}$$

where $x = 2\pi\phi$ and $\bar{\beta} = [A, B, C, D, E, F, G, H, I, J, K, L, M]$. Solve this model with `CURVEFIT`. Overplot the data and the model. Use `XYOUTS` to display the values of n, χ_0^2 and $P(\chi^2 > \chi_0^2)$.

∼ • ∼ Exercise ends here. ∼ • ∼

- Aim of Exercise 9: To understand the limitations of any period finding method.

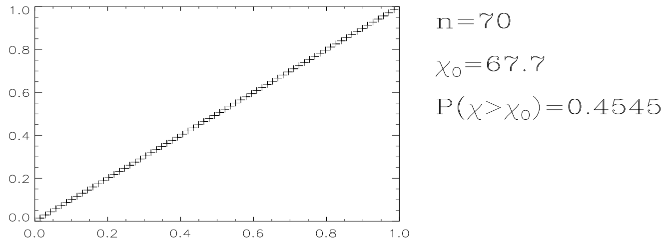

Fig. 17. This figure was made with MODEL7.PRO.

```
; ----------------------------------------------------------------;
; MODEL7.PRO
; ----------------------------------------------------------------;
; An example of how to manipulate numbers to strings and display
; the results.
; ----------------------------------------------------------------;
ERASE & PI1=1.    & IF (PI1 EQ 0.) THEN GOTO,NOPS1 & SET_PLOT,'PS'
DEVICE,/LANDSCAPE & DEVICE,FILENAME='MODEL7.PS'   & NOPS1:
; ----------------------------------------------------------------;
!X.STYLE=1    & !Y.STYLE=1         ; Do not modify SET_XY-limits
X1=0. & X2=1. & Y1=0. & Y2=1.      ; Choose those limits
SET_XY,X1,X2,Y1,Y2                 ; Fix those limits
SET_VIEWPORT,0.1,0.5,0.1,0.5       ; Location on screen
A=FINDGEN(70)/69. & B=A            ; Invent some data to plot
PLOT,A,B,PSYM=1                    ; Plot the data
; ----------------------------------------------------------------;
N=70. & NU=N-3 & CHI=67.66         ; Display this information
PROB=1.DO-CHISQR_PDF(CHI,NU)       ; -"-
TX='!6n=               !3'         ; From hereafter explained
STRPUT,TX,STRING(N,'(I3)'),6       ; during lectures .....
TX=STRCOMPRESS(TX,/REMOVE_ALL)
XYOUTS,X1+1.1*(X2-X1),Y1+0.9*(Y2-Y1),TX,SIZE=2
TX='!4v!D0!N!6=             !3'
STRPUT,TX,STRING(CHI,'(F8.1)'),12
TX=STRCOMPRESS(TX,/REMOVE_ALL)
XYOUTS,X1+1.1*(X2-X1),Y1+0.7*(Y2-Y1),TX,SIZE=2
TX='!6P(!4v>v!D0!N)!6=             !3'
STRPUT,TX,STRING(PROB,'(F7.4)'),20
TX=STRCOMPRESS(TX,/REMOVE_ALL)
XYOUTS,X1+1.1*(X2-X1),Y1+0.5*(Y2-Y1),TX,SIZE=2
; -------------------------------------------------------------------;
IF (PI1 EQ 0.) THEN GOTO,NOPS2 & DEVICE,/CLOSE & SET_PLOT,'X' & NOPS2:
; -------------------------------------------------------------------;
END
; -------------------------------------------------------------------;
```

13. Error estimates for $\bar{\beta}$

- Analytical σ_{β_i} estimates for linear models are relatively easy.
- Analystical σ_{β_i} estimates for nonlinear models are more complicated.
- Analytical $\sigma_{\bar{\beta}}$ estimates rely on concepts like:
  - $\odot$ χ^2–statistics, e.g. χ^2 changes with ± 1
  - $\odot$ Gradient vector $\bar{\gamma}$ determined by $\frac{\partial \chi^2}{\partial \beta_i}$
  - $\odot$ Curvature matrix $[\alpha]$ determined by $\frac{\partial^2 \chi^2}{\partial \beta_i \partial \beta_j}$
  - $\odot$ Covariance matrix $[C] = [\alpha]^{-1}$
  - $\odot$ And ... and ...
- `YFIT=CURVEFIT(X,Y,W,B,EB,FUNCTION_NAME='IDEA')` gives `EB` $= \sigma_{\bar{\beta}}$.
- There are (at least) five general limitations for obtaining reliable $\sigma_{\bar{\beta}}$ estimates:
  - $\odot$ 1: The accuracy $\bar{\sigma}$ of the data must be known.
  - $\odot$ 2: The $\bar{\sigma}$ distribution must be gaussian.
  - $\odot$ 3 Outliers have been removed before modelling. For example, erroneous data or (even correct) data (like flares) are not modelled.
  - $\odot$ 4: The model is reasonable. For example, if you fit the line $g = at + b$ to sinusoidal variation, then $\sigma_{\bar{\beta}}$ is meaningless.
  - $\odot$ 5: The dimensions of the model partial derivatives are correct, especially in nonlinear models. For example, if the nonlinear model is $g = M + A\cos(t/P) + B\sin(t/P)$ and $\bar{\beta} = [M, A, B, P]$, then $\frac{\partial g}{\partial P} = (t/P^2)[A\sin(t/P) - B\cos(t/P)]$. The correct units for the multiplier t/P^2 are $t = t - t_1$, where t_1 is the time of the first modelled observation $y_1 = y(t_1)$.
- Numerical $\sigma_{\bar{\beta}}$ estimates obtained with Bootstrap (uses $\bar{\epsilon}$) or Monte Carlo (uses $\bar{\sigma}$).
- We avoid analytical problems by combining CURVEFIT and bootstrap.
- Basic principle: "Avoid black boxes"

14. Other model parameters

- Other model parameters may, or may not, depend on free parameters $\bar{\beta}$, e.g. amplitude $A = \max[g] - \min[g]$.
  - ⊙ We will later solve, e.g. A, $t_{\mathrm{min},1}$, ... $t_{\mathrm{min,K}}$.
- Good news: If $\beta_{\mathrm{i}} \pm \sigma_{\beta_{\mathrm{i}}}$ available, the error estimates for other model parameters can be derived with the same principle for linear and nonlinear models.
- Not so good news: Analytical error estimates for other model parameters may be complicated, or impossible. But bootstrap can give numerical solutions even for these cases.
- Error estimates for $f = f(\bar{\beta})$ with known σ_{β_i} $(1 \leq i \leq M)$, if the free parameters are uncorrelated (usually they are not!):

$$\sigma_{\mathrm{f}}^2 = \sum_{\mathrm{i=1}}^{\mathrm{M}} \left[\frac{\partial f}{\partial \beta_{\mathrm{i}}} \sigma_{\beta_{\mathrm{i}}} \right]^2$$

- **Examples**
  - ⊙ $A \pm \sigma_{\mathrm{A}}$ and $B \pm \sigma_{\mathrm{B}}$ known. What is the error for $C = A + B$?

$\frac{\partial C}{\partial A} = 1$, $\frac{\partial C}{\partial B} = 1 \Rightarrow \sigma_{\mathrm{C}}^2 = 1\sigma_{\mathrm{A}}^2 + 1\sigma_{\mathrm{B}}^2 \Leftrightarrow \sigma_{\mathrm{C}} = \sqrt{\sigma_{\mathrm{A}}^2 + \sigma_{\mathrm{B}}^2}$

  - ⊙ $g(x) = Ax + B$, $\bar{\beta} = [A, B]$, $\bar{\sigma}_{\beta} = [\sigma_{\mathrm{A}}, \sigma_{\mathrm{B}}]$. What is the error for $g(x)$?

$\frac{\partial g(x)}{\partial A} = x$, $\frac{\partial g(x)}{\partial B} = 1 \Rightarrow \sigma_{\mathrm{g(x)}}^2 = x^2 \sigma_{\mathrm{A}}^2 + 1\sigma_{\mathrm{B}}^2 \Leftrightarrow \sigma_{\mathrm{g(x)}} = \sqrt{(x\sigma_{\mathrm{A}})^2 + \sigma_{\mathrm{B}}^2}$.

Note that $g(x)$ error depends on x.

15. Exercises of analytical solutions for the "other" model parameters and their errors

- These two voluntary exercises will later motivate your appreciation of bootstrap.
- You may assume that the free parameters are uncorrelated in both exercises.
- **Voluntary exercise** The model

  $g(x) = M + B_1 \cos x + C_1 \sin x$

  has $\bar{\beta} \pm \bar{\sigma}_\beta = [M, B_1, C_1] \pm [\sigma_{\mathrm{M}}, \sigma_{\mathrm{B_1}}, \sigma_{\mathrm{C_1}}]$

  and $0 \leq x < 2\pi$.

  1a) Solve $A = \max[g] - \min[g]$ and σ_{A}.

  1b) The primary minimum $x_{\min}$ fulfils $g(x_{\min}) = \min[g]$. **Solve** $x_{\min} \pm \sigma_{\mathrm{x_{min}}}$.

- **Voluntary exercise** The model

  $g(x) = M + B_1 \cos x + C_1 \sin x + B_2 \cos(2x) + C_2 \sin(2x)$,

  has $\bar{\beta} \pm \bar{\sigma}_\beta = [M, B_1, B_2, C_1, C_2] \pm [\sigma_{\mathrm{M}}, \sigma_{\mathrm{B_1}}, \sigma_{\mathrm{B_2}}, \sigma_{\mathrm{C_1}}, \sigma_{\mathrm{C_2}}]$

  and $0 \leq x < 2\pi$.

  You may assume that the free parameters are uncorrelated.

  2a) Solve $A = \max[g] - \min[g]$ and σ_{A}.

  2b) This $g(x)$ may have one or two minima, i.e. $x_{\min,1}$ (primary) and $x_{\min,2}$ (secondary). **Solve** $x_{\min,1} \pm \sigma_{\mathrm{x_{min,1}}}$ and $x_{\min,2} \pm \sigma_{\mathrm{x_{min,2}}}$.

  <u>Warning:</u> Just give up, if it takes more than 1^{h} to solve 2ab). <u>Why</u> is it so complicated?

16. Grid search (i.e . GSch)

- PSch detected a frequency f' $\Rightarrow$ GSch determines a more accurate value around f'.
- PSch tested integer multiples of $\Delta f_{\mathrm{pilot}} = [OFAC\ D_{\max}]^{-1}$, where suitable values were $OFAC \approx 10$.
- GSch tests integer multiples of $\Delta f_{\mathrm{grid}} = [OFAC\ \Delta T]^{-1}$, where $\Delta T = t_{\max} - t_{\min} = t_{\mathrm{n}} - t_1$ and suitable values are $OFAC \approx 10$.
- GSch tests the frequency interval $f' \pm 5\Delta f_{\mathrm{pilot}}$.
- In realistic cases $D_{\max} \leq \Delta T \Rightarrow \Delta f_{\mathrm{grid}} \leq \Delta f_{\mathrm{pilot}} \Rightarrow$ The tested frequency grid in GSch is denser than in PSch.
- PSch "model" was τ.

  ⊙ For example, $\tau = 0.25$ could detect a box function or a sinusoid. The PSch "model" was therefore not strictly fixed.
- GSch model is strictly fixed to

  $$g(\bar{\beta}) = g(t, w, \bar{\beta}) = M + \sum_{\mathrm{k}=1}^{\mathrm{K}} B_{\mathrm{k}} \cos(k2\pi f t) + C_{\mathrm{k}} \sin(k2\pi f t),$$

  where $\bar{\beta} = [M, B_1, ..., B_{\mathrm{K}}, C_1, ..., C_{\mathrm{K}}]$, $K = [4\tau]^{-1}$.
- Because the tested frequency f is not a free parameter, this model is <u>linear</u> and the solution for $\bar{\beta}$ is unique.
- A logical connection between PSch and GSch "models" is $\tau = 1/(4K)$
- PSch did not correlate <u>all data</u>, because the functions $W(t_{\mathrm{i,j}})$ and $Z(\phi_{\mathrm{f,i,j}})$ excluded some data points $y_{\mathrm{i,j}}$.
- The GSch correlates <u>all data</u>. The GSch periodogram

  $$\Theta_{\mathrm{grid}}(f) = 2 \left[\sum_{\mathrm{i}=1}^{\mathrm{n}} w_{\mathrm{i}} \right]^{-1} \sum_{\mathrm{i}=1}^{\mathrm{n}} w_{\mathrm{i}} \left[y(t_{\mathrm{i}}) - g(t_{\mathrm{i}}, w_{\mathrm{i}}, \bar{\beta}_{\mathrm{f}}) \right]^2,$$

  is based on a linear LSF performed for <u>all data</u> at each tested f. The LSF solves $\bar{\beta}_{\mathrm{f}} = [M, B_1, ..., B_{\mathrm{K}}, C_1, ..., C_{\mathrm{K}}]$.
- <u>Note:</u> The latter sum is $\chi^2(\bar{\beta}_{\mathrm{f}})$, but $\Theta_{\mathrm{grid}}(f) \neq \chi^2(\bar{\beta}_{\mathrm{f}})$.
- GSch performs a linear LSF at each tested f. The GSch periodogram is

  $$\Theta_{\mathrm{grid}}(f) = \frac{2\chi^2(\bar{\beta}_{\mathrm{f}})}{\sum_{\mathrm{i}=1}^{\mathrm{n}} w_{\mathrm{i}}} = C\chi^2(\bar{\beta}_{\mathrm{f}}),$$

  where $C = 2\left[\sum_{\mathrm{i}=1}^{\mathrm{n}} w_{\mathrm{i}})\right]^{-1}$ has the same constant value at every tested f.

- Logical GSch program structure is given below.
- Note that the units of time must be $t_i = t_i - t_1$!

	Task	Math and/or programming
1	Subroutine	`FUNCT` for `CURVEFIT`, $g(\bar{\beta})$ and $\partial g/\partial\beta_{\mathrm{i}}$
2	Input	$t_{\mathrm{i}} = t_{\mathrm{i}} - t_1$, y_{i}, σ_{i}, K, f', $\Delta f_{\mathrm{pilot}} = 1/D_{\mathrm{max}}, OFAC$
3	Output	f, $\Theta_{\mathrm{grid}}(f)$, f_{best} at $\Theta_{\mathrm{grid}}(f = f_{\mathrm{best}})$ minimum and $\bar{\beta}_{\mathrm{f}_1}$
4	Derive all tested f	All integer multiples of Δf_{grid} between $[f' - 5\Delta f_{\mathrm{pilot}}, f' + 5\Delta f_{\mathrm{pilot}}]$ Create a vector Θ_{grid} of same length.
5	Begin f loop	
6	Compute $\Theta_{\mathrm{grid}}(f)$	$\Theta_{\mathrm{grid}}(f) = 2\left[\sum_{\mathrm{i}=1}^{n} w_{\mathrm{i}}\right]^{-1} \chi^2(\bar{\beta}_{\mathrm{f}})$, where `CURVEFIT` determines $\bar{\beta}_{\mathrm{f}_{\mathrm{best}}}$
7	End f loop	
8	Interpret $\Theta_{\mathrm{grid}}(f)$	Graphical presentation, identify f_{best} and derive/store $\bar{\beta}_{\mathrm{f}_{\mathrm{best}}}$.

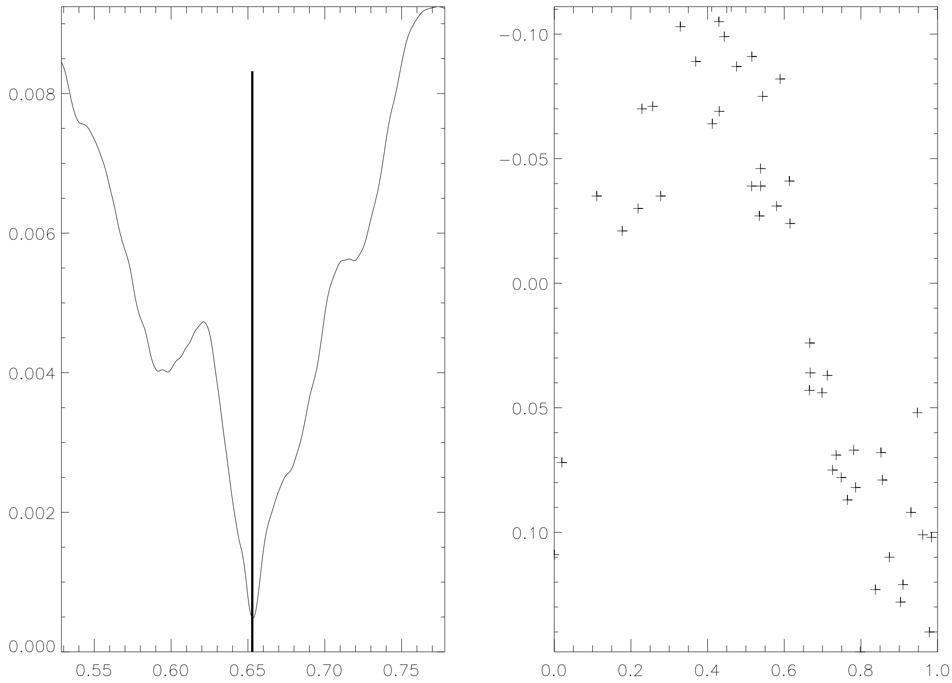

Fig. 18. One example of a solution for **Exercise 10.**

- **Exercise 10.** Edit an IDL-program EXERCISE10.PRO that calculates the GSch periodogram for your data of Exercise 4. Use the best frequency f' detected in Exercise 4 as the mid point of your test. Use $D_{\max} = 40$ days to calculate $\Delta f_{\text{pilot}} = 1/D_{\max}$ Plot the periodogram $\Theta_{\text{grid}}(f)$ within the frequency interval $f' \pm 5\Delta f_{\text{pilot}}$ and indicate the best frequency f_{best} with a vertical line. Use $OFAC = 10$ in computing the tested frequencies. Because these data have no error estimates, fix the weights to unity, i.e. $w_i = 1$. Plot the data as a function of phase calculated with the best frequency. One example of a solution is given in Fig. 18

$\sim \bullet \sim$ Exercise ends here. $\sim \bullet \sim$

17. Writing into a file in IDL

- The analysis results must sooner or later be stored somewhere.
- Here is an example of, how to write into a file in IDL.

```
; ----------------------------------------------------------------
; MODEL8.PRO:
; ===========
; -An example of writing into a file in IDL
TEXT='Whatever'
T0=6543.56789          & T=T0+FINDGEN(100)+0.2*(RANDOMU(S,100)-0.5)
A=0.2                  & SN=5.  & E=(A/SN)*RANDOMN(S,100)
P=15.7119D0            & Y=A*SIN(2.*!PI*T/P)+E
E=ABS((A/SN)*RANDOMN(S,100)) ; Why new errors?
PLOTERR,T,Y,E,PSYM=4         ; Plot just to check
; ----------------------------------------------------------------
OPENW,1,'MODEL8A.DAT'
PRINTF,1,"$(A18)",TEXT      ; String with 18 characters
PRINTF,1,"$(F14.3)",T0      ; Float with 14 characters: 3 decimals
FOR i=0,N_ELEMENTS(T)-1 DO BEGIN
PRINTF,1,"$(I5,X,F14.3,2X,F8.3,X,F6.2)",$  ; I=integer, 2X=2 empty
i+1,T(i),Y(i),E(i)
ENDFOR & CLOSE,1
; ---------------------------------------------------------------
END
; ---------------------------------------------------------------
```

- Creates a file having this kind of a beginning

```
          Whatever
      6543.568
    1       6543.474     0.041   0.00
    2       6544.551    -0.040   0.01
    3       6545.578    -0.058   0.09
    4       6546.594    -0.236   0.06
    5       6547.574    -0.236   0.04
    6       6548.544    -0.230   0.03
    7       6549.596    -0.208   0.07
```

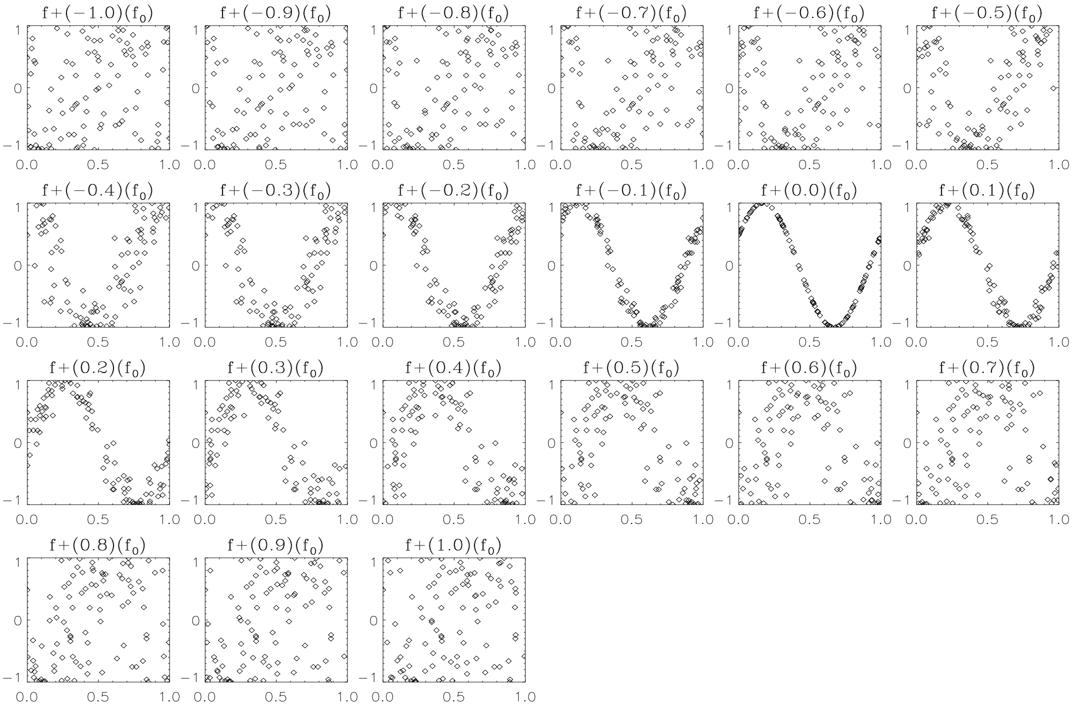

Fig. 19. When does any model with a frequency f stop making sense?

18. Number of independent frequencies

- Problem: How far from any arbitrary P value does a model, like $g(t) = \sin\left(\frac{2\pi t}{P}\right) = \sin 2\pi f t$, stop making sense?
- This is tested with `MODEL9.PRO`. We edit it and test different periods.
- Result: The model stops making sense at about $f \pm f_0$, or even earlier (see Fig. 19)
- $f\Delta T + (f \pm f_0/2)\Delta T = \pm 1/2 \Leftrightarrow$ If f changes with $\pm f_0/2$, the number of integer P during ΔT changes with $\pm 1/2 \Leftrightarrow$ Phases change by 1, and are totally rearranged.
- Interpretation 1: Were we searching for this frequency f, the periodogram values (e.g. $\Theta_{\rm pilot}(f)$, $\Theta_{\rm grid}(f)$) would become independent of each other within a range of $f \pm f_0/2$.
- Interpretation 2: Since it does not matter what frequency is tested, the number of statistically independent tests between $f_{\max}$ and $f_{\min}$ is simply:

  $$m = \mathrm{INT}[(f_{\max} - f_{\min})/f_0]$$

  where INT removes the decimal part of $(f_{\max} - f_{\min})/f_0$, e.g. INT[6.234] = 6.
- Interpretation 3: It does not matter how many frequencies are tested within a range of f_0, because the periodogram values will correlate within $f \pm f_0/2 \Rightarrow$ A periodogram will not undergo "sudden" changes within $f \pm f_0/2$.
- Note: This was already tested in Exercise 2.!

```
; ----------------------------------------------------------------------
; MODEL9.PRO
; ==========
; Distance between frequencies "when model stops making sense" .....
; ----------------------------------------------------------------------
P=1.0D0 & READ,'Give the period = ',P        ; Give any period P!
F=1.D0/P  & DT=30*P                          ; f and Delta T
T=DT*RANDOMN(S,100) & Y=SIN(2.D0*!PI*F*T)    ; Random t & sinusoid
F0=1.D0/(MAX(T)-MIN(T))                      ; f_0
C=21.D0  ; Frequency interval [f-f_0,f+f_0] divided into C parts.
C1=1.D0*FIX(C/2.)                            ; +/- both sides f_0
F_STEP=F0/(C1)                               ; Steps in f_0 units
; ---------------------------------------------------------------------
L=0.15 & K=0.22                & W1=.01+FINDGEN(6)*L
W1=[W1,W1,W1,W1,W1]            & W2=W1+.8*L
W4=.97-FIX(FINDGEN(25)/6.D0)*K & W3=W4-.7*K
; ---------------------------------------------------------------------
!P.NOERASE=1      & !X.TICKS=2 & !Y.TICKS=2 & !P.CHARSIZE=0.8
ERASE & PI1=1.    & IF (PI1 EQ 0.) THEN GOTO,NOPS1 & SET_PLOT,'PS'
DEVICE,/LANDSCAPE & DEVICE,FILENAME='MODEL9.PS'   & NOPS1:
; ---------------------------------------------------------------------
FOR D=0,C-1 DO BEGIN
SET_VIEWPORT,W1(D),W2(D),W3(D),W4(D)
QQ=1.D0*(D-C1)    & F_NOW=F+(D-C1)*F_STEP
; -It is always good to check what is done! _________________; Check
NY1=STRING(QQ/C1,'(F4.1)')   & TX='!6f+(       )(f!D0!N)!3'   ; Check
STRPUT,TX,NY1,5   & !P.TITLE=STRCOMPRESS(TX,/REMOVE_ALL)      ; Check
PRINT,'Round=',QQ,',Shift in f_0',(F_NOW-F)/F0,'Title=',TX    ; Check
; ------------------------------------------------------------; Check
PHI=((T-MIN(T))*F_NOW)       & PHI=PHI-FIX(PHI)
PLOT,PHI,Y,PSYM=4,SYMSIZE=0.5
ENDFOR
; ----------------------------------------------------------------------
IF (PI1 EQ 0.) THEN GOTO,NOPS2 & DEVICE,/CLOSE & SET_PLOT,'X' & NOPS2:
; ----------------------------------------------------------------------
END
; ----------------------------------------------------------------------;
```

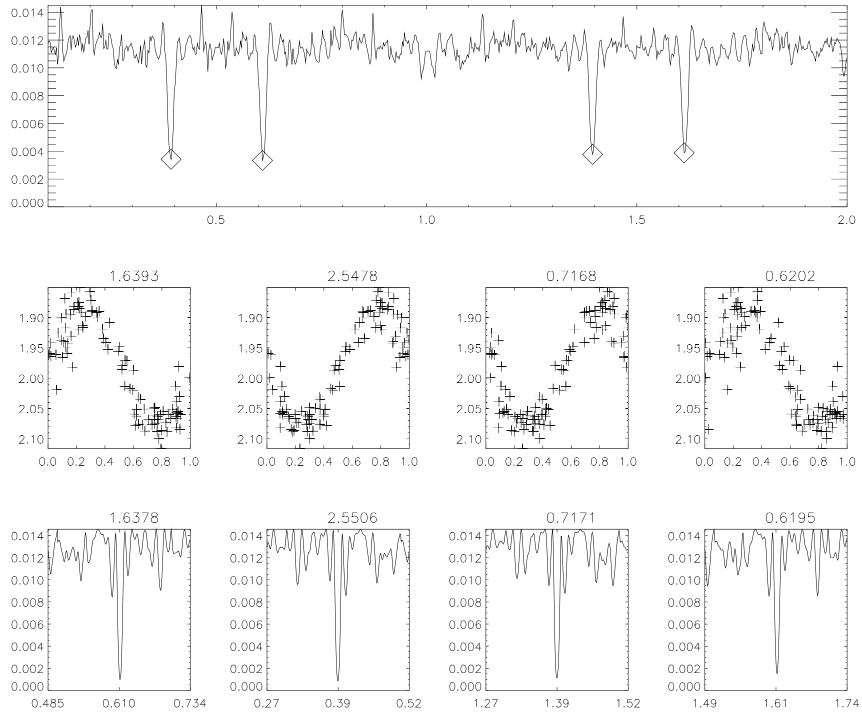

Fig. 20. One example of a solution for **Exercise 11.**

- **Exercise 11.** You will receive a new datafile `STUDENT*_AIJA.DAT` which also contains the errors of the data ($\sigma_{\rm i}$) on the third column. Edit an IDL-program EXERCISE11.PRO that performs combined PSch and GSch analysis of these data for the case $K = 1$. The tested PSch period interval is $P_{\rm min} = 0.5$ and $P_{\rm max} = 10$. Use the flow chart in the home page, i.e. program the subroutines

  `PILOTSEARCH,T,Y,W,PMIN,PMAX,OFAC,K,F,PTHETA,F4,PTHETA4,DMAX`

  and

  `GRIDSEARCH,T,Y,W,F4(i),DMAX,OFAC,F,GTHETA,FBEST,GTHETABEST,BETACUR`

  One example of a solution is given in Fig. 20. The upper plot shows the PSch periodogram. The next four phase plots show the data with the four best periods detected with PSch. The last four plots show the respective GSch periodograms.

$\sim \bullet \sim$ Exercise ends here. $\sim \bullet \sim$

19. Refined search, i.e. RSch

- TSPA model is

$$g(\bar{\beta}) = g(t, \bar{\beta}) = M + \sum_{\mathrm{k}=1}^{\mathrm{K}} B_{\mathrm{k}} \cos(k2\pi f t) + C_{\mathrm{k}} \sin(k2\pi f t),$$

where $\bar{\beta} = [M, B_1, ..., B_{\mathrm{K}}, C_1, ..., C_{\mathrm{K}}, f]$ denotes the free parameters of the model.

- These $Q = K + 2$ free parameters are:
  - ⊙ M = mean
  - ⊙ B_{k} and C_{k} = amplitudes
  - ⊙ f = frequency
- GSch tested a fixed discrete frequency grid
- GSch did **linear least squares fits, because f was not a free parameter**.
- RSch does a **nonlinear least squares fit, because f is a free parameter**.
- The result of this nonlinear Marquardt iteration depends on the trial value for the free parameters $\bar{\beta}_{\mathrm{trial}}$.
- GSch was performed to obtain such a reliable trial value $\bar{\beta}_{\mathrm{trial}} = [\bar{\beta}_{\mathrm{best}}, f_{\mathrm{best}}]$
- RSch begins from $\bar{\beta}_{\mathrm{trial}}$ and performs an iteration that is **continuous** in f.
- RSch gives the final and best values of free parameters, i.e. $\bar{\beta}_{\mathrm{final}}$
- An example of the use of `CURVEFIT` for the case $K = 1$ is given on the next page.
- The flowchart between GSch and RSch is given on the homepage.

`CURVEFIT` in GSch and RSch for $K = 1$

- A suitable GSch subroutine for `CURVEFIT`

```
PRO FUNCT1,X,A,F,PDER
; -The partial derivatives for CURVEFIT in GSch
; -Note that X = 2 Pi f t, i.e. fit in time!
; ----------------------------------------------------------------
F=0.D0+A(0)+A(1)*COS(X)+A(2)*SIN(X) & PDER=DBLARR(N_ELEMENTS(X),3)
PDER(*,0)=1.D0 & PDER(*,1)=COS(X)   & PDER(*,2)=SIN(X)
; ----------------------------------------------------------------
RETURN & END
```

- A suitable RSch subroutine for `CURVEFIT`

```
PRO GUNCT1,X,A,F,PDER
; ----------------------------------------------------------------
; -The partial derivatives for CURVEFIT in Rsch for K=1.
; -Note that X = 2 Pi t, i.e. fit in time!
; ----------------------------------------------------------------
F=0.D0+A(0)+A(1)*COS(A(3)*X)+A(2)*SIN(A(3)*X)
PDER=DBLARR(N_ELEMENTS(X),4)
PDER(*,0)=1.D0 & PDER(*,1)=COS(A(3)*X)   & PDER(*,2)=SIN(A(3)*X)
PDER(*,3)=X*(A(2)*COS(A(3)*X)-A(1)*SIN(A(3)*X))
; ----------------------------------------------------------------
RETURN & END
```

- The value of $\bar{\beta}_{\mathrm{trial}}$, i.e. `BETACUR` derived with the GSch is

```
X=2.D0*!PI*(T-MIN(T))*FBEST
YFIT=CURVEFIT(X,Y,W,A,EA,FUNCTION_NAME='GUNCT1')
BETACUR=[A,FBEST] ; Trial solution for RSch
```

- Note that f is not a free parameter and $t_i = t_i - t_1$.
- The final value of $\bar{\beta}_{\mathrm{final}}$, i.e. `BETAFINAL` derived with the RSch is

```
BETAFINAL=BETACUR
X=2.D0*!PI*(T-MIN(T))
YFIT=CURVEFIT(X,Y,W,BETAFINAL,EA,FUNCTION_NAME='GUNCT1')
```

- Note that f is a free parameter and $t_i = t_i - t_1$.

20. Bootstrap

- What have we learned so far?

  ⊙ Analytical estimates for $\bar{\beta}$ and the errors $\bar{\sigma}_\beta$ are difficult for <u>nonlinear</u> models.

  ⊙ Analytical estimates for $\bar{\beta}$ and the errors $\bar{\sigma}_\beta$ are impossible for both linear and nonlinear model, if the errors for the data ($\bar{\sigma}$) are <u>unknown</u>.

  ⊙ These problems are even more pronounced for other model parameters, e.g. the total amplitude (A) of $\bar{g}$, the epochs of the minima/maxima of $\bar{g}$, etc..

- Bootstrap solves all these problems <u>numerically!</u>

The six stages of bootstrap

- 1: Minimizing χ^2 for $\bar{y}$ with $\bar{w}$ gives the "empirical distribution of residuals"

  $\epsilon_{\rm i} = y(t_{\rm i}) - g(t_{\rm i}, \bar{\beta}_{\rm min}) = y_{\rm i} - g_{\rm i},$

- 2. A random sample $\bar{\epsilon}^*$ is selected from $\bar{\epsilon}$. The number of the *same* $\epsilon_{\rm i}$ entering into $\bar{\epsilon}^*$ may vary between 0 and n in this random sample with replacement. The $\bar{\epsilon}^*$ determine a unique random sample $\bar{w}^*$, where the connection of $w_{\rm i}$ being the weight of $\epsilon_{\rm i}$ is preserved.
- 3. A random sample

  $\bar{y}^* = \bar{g} + \bar{\epsilon}^*$

  is obtained.
- 4. Minimizing χ^2 for $\bar{y}^*$ with $\bar{w}^*$ gives one estimate for $\bar{\beta}'_{\rm min}$, as well as for the "other" parameters of higher order models ($K \geq 2$).

  ⊙ For example, measuring the difference between the minimum and maximum of $\bar{g}(\bar{\beta}'_{\rm min})$ gives a *numerical* A estimate.
- 5. The bootstrap returns to the 2nd stage, until S estimates of $\bar{\beta}'_{\rm min}$ and "other" parameters have been obtained.
- 6. The expectation value and variance for any $\bar{\beta}_{\rm min}$ component are the mean and variance of its S estimates in $\bar{\beta}'_{\rm min}$. The same applies to the S estimates of "other" model parameters, like total amplitude A
- <u>Note:</u>

  ⊙ $\bar{y}$, $\bar{w}$, $\bar{\epsilon}$ and $\bar{g}$ do not change during bootstrap.

  ⊙ $\bar{y}^*$, $\bar{w}^*$ and $\bar{\epsilon}^*$ change during every bootstrap round.

  ⊙ $\bar{\epsilon}^*$ determines $\bar{y}^*$ and $\bar{w}^*$.
- `MODEL10.PRO` gives an example of bootstrap, where the epoch of the minimum (`RESULT1`) and total amplitude (`RESULT2`) are solved numerically.

```
; -------------------------------------------------------------;
;                     MODEL10.PRO
; -Bootstrap example, where minimum and total amplitude solved
; numerically.
; -------------------------------------------------------------;
PRO FU,X,A,F,PDER
F=0.+A(0)+A(1)*COS(X)+A(2)*SIN(X)  & PDER=DBLARR(N_ELEMENTS(X),3)
PDER(*,0)=1.D0  & PDER(*,1)=COS(X) & PDER(*,2)=SIN(X)
RETURN & END
; -------------------------------------------------------------;
PRO MITTAA,XX,GG,AMP,TMIN
; - Solves the total amplitude and first minimum
J=WHERE((GG LT SHIFT(GG,1)) AND (GG LT SHIFT(GG,-1))$
AND     (XX NE MIN(XX))     AND (XX NE MAX(XX))    )
TMIN=MIN(XX(J(0)))                & AMP=MAX(GG)-MIN(GG)
RETURN & END
; -------------------------------------------------------------;
; -Simulated data
N =10.D0 & READ,'Give ..... n = ',N
Z =10.D0 & READ,'Give A/sigma = ',Z
M =10.D0 & READ,'Give ..... M = ',M
B_1=1.D0 & READ,'Give ... B_1 = ',B_1
C_1=1.D0 & READ,'Give ... C_1 = ',C_1
T =2.D0*RANDOMU(S,N)         & T =2.D0*!PI*T
Y =M+B_1*COS(T)+C_1*SIN(T)   & A =MAX(Y)-MIN(Y)
E =(A/Z)*RANDOMN(S,N)        & Y=Y+E   & E =(A/Z)*RANDOMN(S,N)
J =SORT(T) & T=T(J) & Y=Y(J) & E=E(J) ; Sorted in time
; -------------------------------------------------------------;
!P.NOERASE=1  & !X.TICKS=2. & !Y.TICKS=2. & !P.SYMSIZE=1.5
!X.STYLE=1    & !Y.STYLE=1  & !P.CHARSIZE=1.5
; -------------------------------------------------------------;
ROUNDS=10.D0 & READ,'Bootstrap samples     = ',ROUNDS
AIKAA =5.D0  & READ,'Seconds between plots = ',AIKAA
W=E^(-2.D0)  & BETA=DBLARR(3)
G=CURVEFIT(T,Y,W,BETA,SIGMAA,FUNCTION_NAME='FU') & EPSILON=Y-G
; -------------------------------------------------------------;
X1=MIN(T) & X2=MAX(T) & Y1=0.9*MIN(Y-E) & Y2=1.1*MAX(Y+E)
XX=FINDGEN(10001)/10000.D0 & XX=0.D0+X1+(X2-X1)*XX
G_ORIG=BETA(0)+BETA(1)*COS(XX)+BETA(2)*SIN(XX)
MITTAA,XX,G_ORIG,AMP_ORIG,MIN_ORIG
RESULT1=AMP_ORIG & RESULT2=MIN_ORIG
; -------------------------------------------------------------;
FOR QQ=0,ROUNDS-1 DO BEGIN & ERASE ; Bootstrap begins
; -------------------------------------------------------------;
!P.TITLE='!6y!Di!N!3'             ; Data and model
SET_XY,X1,X2,Y1,Y2    & SET_VIEWPORT,.1,0.3,0.7,0.9
PLOTERR,T,Y,E,PSYM=4  & PLOT,XX,G_ORIG,PSYM=0
OPLOT,[X1,X2],MAX(G_ORIG)*[1,1],PSYM=0,LINESTYLE=1
OPLOT,[X1,X2],MIN(G_ORIG)*[1,1],PSYM=0,LINESTYLE=1
OPLOT,MIN_ORIG*[1,1],[Y1,Y2],   PSYM=0,LINESTYLE=2
; ---------------------------------------------------------;
!P.TITLE='!6g!Di!N!3'             ; Model = g
SET_XY,X1,X2,Y1,Y2    & SET_VIEWPORT,0.1,0.3,0.4,0.6
PLOT,T,G,PSYM=4
; ---------------------------------------------------------;
!P.TITLE='!7e!6!Di!N!3'           ; t and epsilon
```

```
Y3=-1.D0*MAX(ABS(EPSILON)) & Y4=-1.D*Y3
SET_XY,X1,X2,Y3,Y4    & SET_VIEWPORT,.1,.3,0.1,0.3
PLOT,T,EPSILON,PSYM=4
; ------------------------------------------------------------;
!P.TITLE='!7e!6!Di!N!U*!N!3'       ; t and epsilon^*
SET_XY,X1,X2,Y3,Y4    & SET_VIEWPORT,0.4,0.6,0.7,0.9
K=FIX(N*RANDOMU(S,N)) ; Random integers  within [0,K-1]
EPSILON_STAR=EPSILON(K) & E_STAR=E(K)
W_STAR=E_STAR^(-2.0D0)  & Y_STAR=G+EPSILON_STAR ; w^* and y^*
PLOT,T,EPSILON_STAR,PSYM=4
; ------------------------------------------------------------;
!P.TITLE='!6y!Di!N!U*!N=g!Di!N+!7e!6!Di!N!U*!N!3' ; t and y^*
SET_XY,X1,X2,Y1,Y2    & SET_VIEWPORT,.4,.6,0.4,0.6
PLOTERR,T,Y_STAR,E_STAR,PSYM=4 & Q=DBLARR(3)
NY=CURVEFIT(T,Y_STAR,W_STAR,Q,SIGMAA,FUNCTION_NAME='FU')
G_NOW=Q(0)+Q(1)*COS(XX)+Q(2)*SIN(XX)
PLOT,XX,G_NOW,PSYM=0
MITTAA,XX,G_NOW,AMP_NOW,MIN_NOW
PRINT,'A = ',AMP_NOW,', ... t_min,1 = ',MIN_NOW
RESULT1=[RESULT1,AMP_NOW] & RESULT2=[RESULT2,MIN_NOW]
OPLOT,[X1,X2],MAX(G_NOW)*[1,1],PSYM=0,LINESTYLE=1
OPLOT,[X1,X2],MIN(G_NOW)*[1,1],PSYM=0,LINESTYLE=1
OPLOT,MIN_NOW*[1,1],[Y1,Y2],   PSYM=0,LINESTYLE=2
; -------------------------------------------------------------;
!P.TITLE='!6A estimates !3'       ;  A estimates
SET_VIEWPORT,0.7,0.9,0.7,0.9
Y5=AVG(RESULT1)-3.*STDEV(RESULT1)
Y6=AVG(RESULT1)+3.*STDEV(RESULT1)
SET_XY,0.,ROUNDS,Y5,Y6 & IF (Q GT 1) THEN PLOT,RESULT1,PSYM=4
; --------------------------------------------------------------;
!P.TITLE='!6t!Dmin,1!N estimates!3' ; t_min,1 estimates
SET_VIEWPORT,0.7,0.9,0.4,0.6
Y5=AVG(RESULT2)-3.*STDEV(RESULT2)
Y6=AVG(RESULT2)+3.*STDEV(RESULT2)
SET_XY,0.,ROUNDS,Y5,Y6 & IF (Q GT 1) THEN PLOT,RESULT2,PSYM=4
; --------------------------------------------------------------;
WAIT,AIKAA                        ; Waits AIKAA seconds
ENDFOR                            ; Bootstrap ends
; --------------------------------------------------------------;
PRINT,'     A = ',RESULT1(0),'   +/-',STDEV(RESULT1(1:ROUNDS-1))
PRINT,' t_min = ',RESULT2(0),'   +/-',STDEV(RESULT2(1:ROUNDS-1))
END
; --------------------------------------------------------------;
```

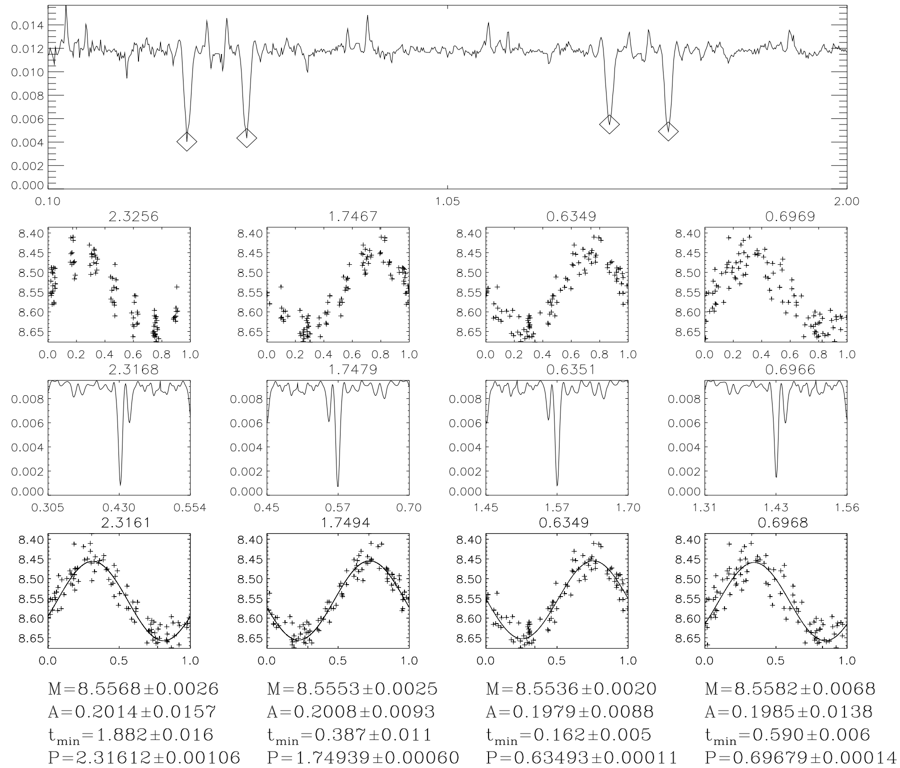

Fig. 21. One example of a solution for **Exercise 12.**

- **Exercise 12.** Rename your program of Exercise 11. Continue your analysis of `STUDENT*_AIJA.DAT` by performing the RSch with $K = 1$. Use the flow chart in the home page, i.e. program the subroutine

  `REFINEDSEARCH,T,Y,W,K,BETACUR,BETAFIN`.

  Solve values and errors of M (mean), A (total amplitude), $t_{\rm min}$ (minimum epoch in time) and P (period) with bootstrap. One example of a solution is given in Fig. 21.

∼ • ∼ Exercise ends here. ∼ • ∼