On misleading solar-climate relationship ...

Supplementary material

Refutal of LeMouel, Kossobokov and Courtillot, JASTP, 2010 (LMKC)

Contact: Bernard Legras, legras@lmd.ens.fr


```
SrcDir = "D:\cygwin\home\legras\h2\empire";
SrcDir = "./";
```

This textbook has been prepared with Mathematica version 6 and 7 and plays with both versions. Earlier versions have not been tested.

Methods

■ Bootstrapping

From Wikipedia: In statistics, bootstrapping is a modern, computer - intensive, general purpose approach to statistical inference, falling within a broader class of resampling methods.Bootstrapping is the practice of estimating properties of an estimator (such as its variance) by measuring those properties when sampling from an approximating distribution.One standard choice for an approximating distribution is the empirical distribution of the observed data.In the case where a set of observations can be assumed to be from an independent and identically distributed population, this can be implemented by constructing a number of resamples of the observed dataset (and of equal size to the observed dataset), each of which is obtained by random sampling with replacement from the original dataset.It may also be used for constructing hypothesis tests.It is often used as an alternative to inference based on parametric assumptions when those assumptions are in doubt, or where parametric inference is impossible or requires very complicated formulas for the calculation of standard errors.The advantage of bootstrapping over analytical methods is its great simplicity - it is straightforward to apply the bootstrap to derive estimates of standard errors and confidence intervals for complex estimators of complex parameters of the distribution, such as percentile points, proportions, odds ratio, and correlation coefficients.The disadvantage of bootstrapping is that while (under some conditions) it is asymptotically consistent, it does not provide general finite - sample guarantees, and has a tendency to be overly optimistic. The apparent simplicity may conceal the fact that important assumptions are being made when undertaking the bootstrap analysis (e.g.independence of samples) where these would be more formally stated in other approaches.

Here we use bootstrapping by resampling over a permutation of full years. This preserves the correlation of temperature over the sequence of the year and allows to test whether the years selected in H and L ensemble exhibit properties different from a random drawing. The number of drawings (10000) is large over the number of years which are considered.

Shuffle resamples a dataset made of temperatures over integer number of 365 years by shuffling randomly the years. The dataset is first partitioned into years by Partition and the resampling is done RandomSample. It is finally flattened b Flatten into a unique list.

```
Shuffle[dd_, Ns_] := Flatten[RandomSample[Partition[dd, 365], Ns]];
```

Student t-test

Two-sided p-value

Student-t test is the standard test to determine whether two means are different. The null hypothesis is that the two means over the H and L subsets are not different.

MeanDifferenceTest[*list*$_1$, *list*$_2$, $\Delta\mu_0$] gives a *p*-value for the test that the difference between the means μ_1 and μ_2 of the populations from which *list*$_1$ and *list*$_2$ were sampled is significantly different from $\Delta\mu_0$.
If the variances for the two populations are assumed equal and unknown, the test is based on Student ' s*t*distribution with Length[*list*$_1$] + Length[*list*$_2$] − 2 degrees of freedom.
If the population variances are not assumed known and not assumed equal, Welch ' s approximation for the degrees of freedom is used.

```
Needs["HypothesisTesting`"]
```

Version with assumed equal variances

```
StudentTest[{List1_, List2_}] := TwoSidedPValue /. MeanDifferenceTest[List1, List2, 0,
   EqualVariances → True, TwoSided → True]
```

Version with variances not assumed to be equal

```
StudentTest2[{List1_, List2_}] := TwoSidedPValue /. MeanDifferenceTest[List1, List2, 0,
   EqualVariances → False, TwoSided → True]
```

Kolmogorov - Smirnov test

Calculation based on Numerical Recipes

$$\text{QKS}[\lambda\_] := 2\sum_{j=1}^{\infty}(-1)^{j-1}\,e^{-2 j^2 \lambda^2}$$

A fast vectorized algorithm (original to our knowledge) is used to calculate the KS distance. It may fail when the two lists contain identical values (not the case in the present application).

```
KolmoSmirnov[{List1_, List2_}] :=
 Module[{N1 = Length[List1], N2 = Length[List2], XX, XXO, DD, Ne},
  XX = Join[Transpose[{List1, ConstantArray[1./N[N1], N1], ConstantArray[0., N1]}],
    Transpose[{List2, ConstantArray[0., N2], ConstantArray[1./N[N2], N2]}]];
  XXO = Transpose[SortBy[XX, First]];
  DD = Max[Abs[Accumulate[XXO[[2]]] - Accumulate[XXO[[3]]]]];
  Ne = (N1 N2)/(N1 + N2);
  Return[QKS[(Sqrt[Ne] + 0.12 + 0.11/Sqrt[Ne]) DD]]];
```

- **Confidence interval for bootstrapping**

```
BinSize = 0.05;
BinVal = Table[-2 + BinSize / 2 + (i - 1) BinSize, {i, 1, 4 / BinSize}];
Pdf = BinCounts[#, {-2, 2, BinSize}] / Length[#] / BinSize &;

Confidence[Refpdf_, bnd_] := Module[{GetFirst, GetLast, First, Last},
  GetFirst = Flatten[
     Select[Transpose[{BinVal, Accumulate[#] * BinSize}], Function[x, x[[2]] > bnd], 1]] &;
  GetLast = Flatten[Select[Transpose[{BinVal, Accumulate[#] * BinSize}],
      Function[x, x[[2]] > 1 - bnd], 1]] &;
  First = Transpose[Map[GetFirst, Refpdf]][[1]];
  Last = Transpose[Map[GetLast, Refpdf]][[1]] - BinSize;
  Return[{First, Last}]]
```

- **Composite operations**

```
GatherYear[temp_, {dl_, ll_}] := Take[Drop[temp, dl * 365], ll * 365]
SolarSplit[temp_, years_] := Module[{HighTemp, LowTemp},
  HighTemp = Partition[Flatten[MapThread[GatherYear[temp, #] &, {years[[1]]}]], 365];
  LowTemp = Partition[Flatten[MapThread[GatherYear[temp, #] &, {years[[2]]}]], 365];
  Return[{HighTemp, LowTemp}]]
SolarDiff[temp_, years_] := Mean[#[[1]]] - Mean[#[[2]]] & @@ {SolarSplit[temp, years]}
StdDev[temp_, years_] :=
```

$$\sqrt{\left(\frac{(\text{Length}[\#[[1]]]-1)\ \text{Variance}[\#[[1]]]+(\text{Length}[\#[[2]]]-1)\ \text{Variance}[\#[[2]]]}{\text{Length}[\#[[1]]]+\text{Length}[\#[[2]]]-2}\left(\frac{1}{\text{Length}[\#[[1]]]}+\frac{1}{\text{Length}[\#[[2]]]}\right)\right)}\ \&\ @@\ \{\text{SolarSplit}[\text{temp},\ \text{years}]\}$$

- **Miscellaneous**

```
Needs["PlotLegends`"]

Off[NIntegrate::nlim]
```

Using the series from Praha since 1775

- **Read data and preprocess them**

```
snm = OpenRead[SrcDir <> "TN_nonblend_Praha.datm"];
Skip[snm, String, 14]; (* skip header *)
data = Transpose[ReadList[snm, {Number, Number, Number, Number}]] /. -9999 → Missing[];
(* read data until the end of the file *)
Close[snm];
time = data[[2]];
tempTN = 0.1 * data[[3]] /. x___ Missing[] → Missing[] ;
snm = OpenRead[SrcDir <> "TX_nonblend_Praha.datm"];
Skip[snm, String, 14]; (* skip header *)
data = Transpose[ReadList[snm, {Number, Number, Number, Number}]] /. -9999 → Missing[];
(* read data until the end of the file *)
Close[snm];
tempTX = 0.1 * data[[3]] /. x___ Missing[] → Missing[] ;
year = Quotient[time, 10 000]; time = time - 10 000 year;
month = Quotient[time, 100]; day = time - 100 month;
time = Transpose[{year, month, day}];
```

- **Eliminate Missing data**

Missing data for Praha are grouped at the end of the record and for recent years (beyond 2005)

```
tempTN = Take[tempTN, 84 126];
tempTX = Take[tempTX, 84 126];
time = Take[time, 84 126];
{Length[Cases[tempTN, _Missing]], Length[Cases[tempTX, _Missing]]}
```

```
{0, 0}
```

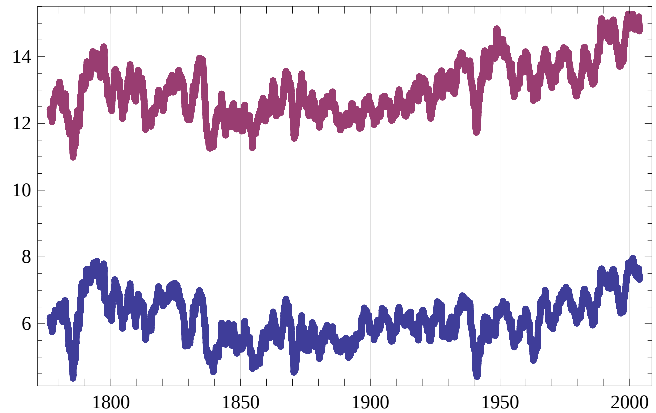
No data is marked as missing after removal of the last years. This helps further processes to perform faster than when Missing[] data are left in the list.

- **3-year filter**

A simple 3 - year moving average is applied to the data, as in LMKC. The time list is truncated accordingly.

```
tempTN3 = MovingAverage[tempTN, 3 * 365];
tempTX3 = MovingAverage[tempTX, 3 * 365];
time3 = Drop[Drop[time, 365 + 182], -365 - 182];
```

```
DateListPlot[{Transpose[{time3, tempTN3}], Transpose[{time3, tempTX3}]}]
```



This is figure 1 of LeMouel et al.

Figure for publication

Calculation on unfiltered data

For commented version of the calculation, see Calculation on 21 - day filtered data below

Calculating and removing annual cycle (on full data)

First get rid of the 29 Febs

```
TT = Evaluate[DeleteCases[Evaluate[Transpose[{tempTN, time}]], {_Real, {_Integer, 2, 29}}]];
{tempTN4, time4} = Transpose[TT];
TT =
  Evaluate[DeleteCases[Evaluate[Transpose[{tempTX, time}]], {_Real, {_Integer, 2, 29}}]];
{tempTX4, time4} = Transpose[TT];
```

Remove incomplete last year

```
tempTN5 = Take[tempTN4, 83 950];
tempTX5 = Take[tempTX4, 83 950];
time5 = Take[time4, 83 950];
{time5[[1]], time5[[-1]]}
```

```
{{1775, 1, 1}, {2004, 12, 31}}
```

Calculate and remove annual cycle

```
tempannual = Map[Mean, Transpose[Partition[tempTN5, 365]]];
TMP = Take[Nest[Join[#, #] &, tempannual, 8], 230 * 365];
tempTN6 = tempTN5 - TMP;
tempannual = Map[Mean, Transpose[Partition[tempTX5, 365]]];
TMP = Take[Nest[Join[#, #] &, tempannual, 8], 230 * 365];
tempTX6 = tempTX5 - TMP;
```

Selecting high and low solar years for unfiltered data

```
SolarCycles = {{3, 1775, 1784}, {4, 1785, 1798}, {5, 1799, 1810}, {6, 1811, 1823},
   {7, 1824, 1833}, {8, 1834, 1843}, {9, 1844, 1856}, {10, 1857, 1867}, {11, 1868, 1878},
   {12, 1879, 1889}, {13, 1890, 1901}, {14, 1902, 1913}, {15, 1914, 1923},
   {16, 1924, 1933}, {17, 1934, 1944}, {18, 1945, 1954}, {19, 1955, 1964},
   {20, 1965, 1976}, {21, 1977, 1986}, {22, 1987, 1996}, {23, 1997, 2004}};
High = {3, 4, 8, 9, 11, 18, 19, 20, 21, 22, 23};
HighYears = {{0, 1798 - 1775 + 1}, {1834 - 1775, 1856 - 1834 + 1},
  {1868 - 1775, 1878 - 1868 + 1}, {1945 - 1775, 2004 - 1945 + 1}}
Low = {5, 6, 10, 12, 13, 14, 15, 16, 17};
LowYears =
 {{1799 - 1775, 1833 - 1799 + 1}, {1857 - 1775, 1867 - 1857 + 1}, {1879 - 1775, 1944 - 1879 + 1}}
Years = {HighYears, LowYears};
a1 = x /. FindRoot[NIntegrate[
     PDF[StudentTDistribution[Length[tempTN5] / 365 - 2], y], {y, 0, x}] == 0.45, {x, 1.6}]
```

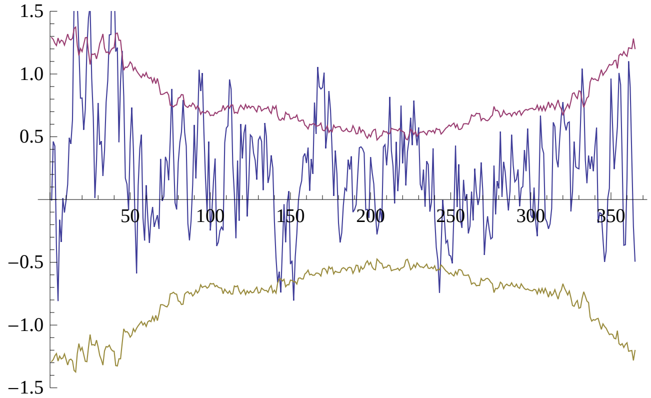
```
{{0, 24}, {59, 23}, {93, 11}, {170, 60}}
```

```
{{24, 35}, {82, 11}, {104, 66}}
```

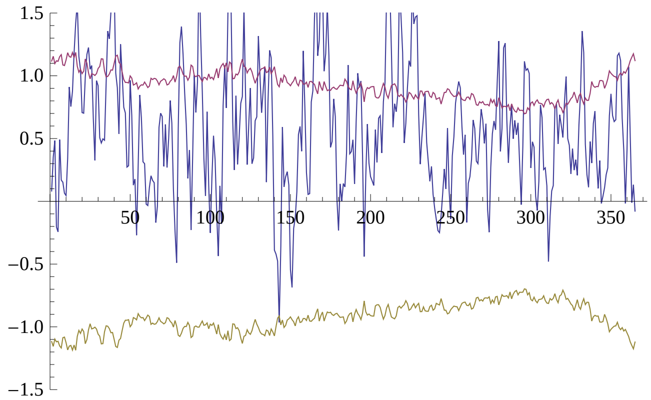
```
1.65156
```

Calculate the composite difference between high and low solar years unfiltered data

```
ListPlot[{SolarDiff[tempTN6, Years], a1 StdDev[tempTN6, Years], -1 a1 StdDev[tempTN6, Years]},
 Joined → True, PlotRange → {Automatic, {-1.5, 1.5}}]
```




```
ListPlot[{SolarDiff[tempTX6, Years], a1 StdDev[tempTX6, Years], -a1 StdDev[tempTX6, Years]},
 Joined → True, PlotRange → {Automatic, {-1.5, 1.5}}]
```



- ### Analysis after removal of the last 5 cycles (ensemble IV according to LeMouel et al.)

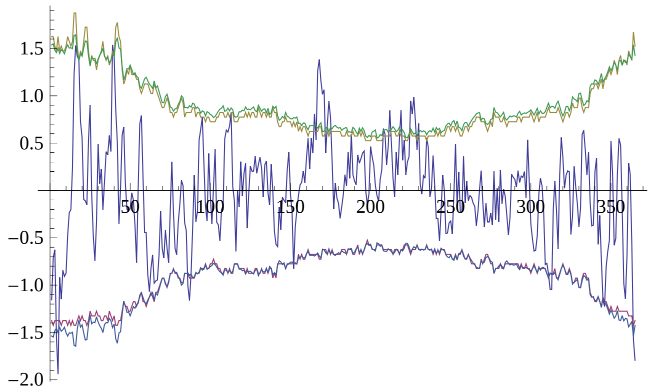
We drop the last 5 cycles or 50 years corresponding to the period where global warming is felt

```
tempTN6IV = Drop[tempTN6, -50 * 365];
tempTX6IV = Drop[tempTX6, -50 * 365];
HighIV = {3, 4, 8, 9, 11, 18};
HighYearsIV = {{0, 1798 - 1775 + 1}, {1834 - 1775, 1856 - 1834 + 1},
   {1868 - 1775, 1878 - 1868 + 1}, {1945 - 1775, 1954 - 1945 + 1}};
YearsIV = {HighYearsIV, LowYears}
RefTNIV = Transpose[Table[SolarDiff[Shuffle[tempTN6IV, 180], YearsIV], {i, 1, 10 000}]];
RefTXIV = Transpose[Table[SolarDiff[Shuffle[tempTX6IV, 180], YearsIV], {i, 1, 10 000}]];
RefpdfTNIV = Map[Pdf, RefTNIV];
RefpdfTXIV = Map[Pdf, RefTXIV];
Sig10TNIV = Confidence[RefpdfTNIV, 0.05];
Sig10TXIV = Confidence[RefpdfTXIV, 0.05];
a1IV = x /. FindRoot[NIntegrate[
      PDF[StudentTDistribution[Length[tempTN6IV] / 365 - 2], y], {y, 0, x}] == 0.45, {x, 1.6}]
```

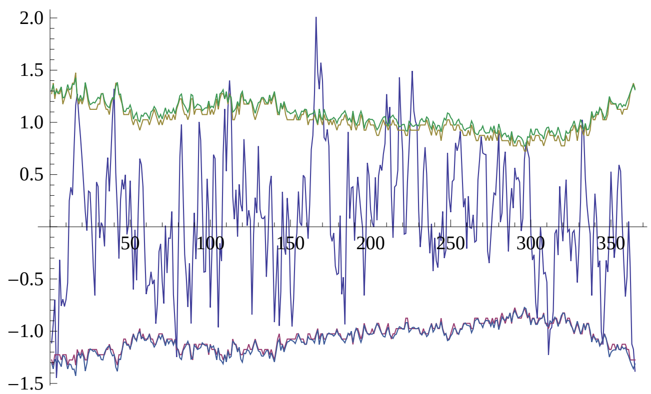
```
{{{0, 24}, {59, 23}, {93, 11}, {170, 10}}, {{24, 35}, {82, 11}, {104, 66}}}
```

```
1.65346
```

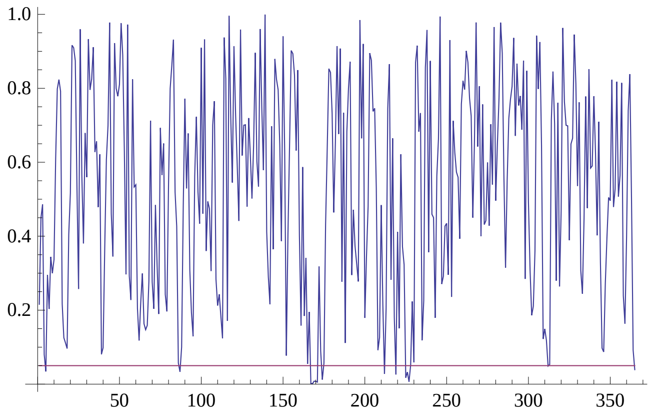
```
ListPlot[{SolarDiff[tempTN6IV, YearsIV], Sig10TNIV[[1]], Sig10TNIV[[2]],
  a1IV StdDev[tempTN6, YearsIV], -a1IV StdDev[tempTN6, YearsIV]}, Joined → True]
```




```
ListPlot[{SolarDiff[tempTX6IV, YearsIV], Sig10TXIV[[1]], Sig10TXIV[[2]],
  a1IV StdDev[tempTX6IV, YearsIV], -a1IV StdDev[tempTX6IV, YearsIV]}, Joined → True]
```



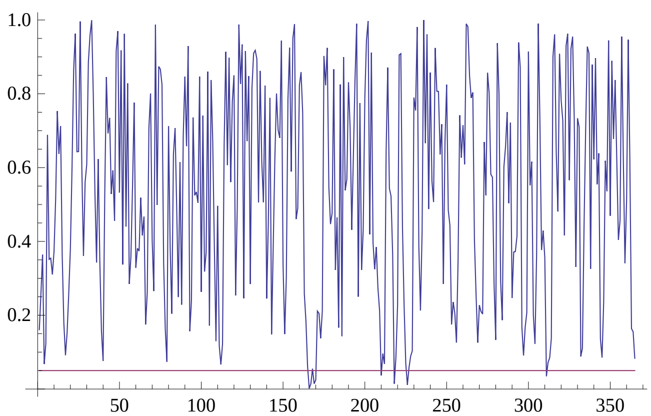
```
PValStudent = StudentTest /@ Transpose[Transpose /@ SolarSplit[tempTN6IV, YearsIV] ];
ListPlot[{PValStudent, ConstantArray[0.05, 365]}, Joined → True]
Mean[PValStudent]
```



```
0.525786
```

$$\text{ListPlot}\left[\text{Transpose}\left[\left\{\frac{1}{365}\text{Range}[365],\ \text{Sort}[\text{PValStudent}] - \frac{1}{365}\text{Range}[365]\right\}\right]\right]$$

```
PValStudent = StudentTest /@ Transpose[Transpose /@ SolarSplit[tempTX6IV, YearsIV] ];
ListPlot[{PValStudent, ConstantArray[0.05, 365]}, Joined → True]
Mean[PValStudent]
```



```
0.533926
```

$$\text{ListPlot}\left[\text{Transpose}\left[\left\{\frac{1}{365}\text{Range}[365],\ \text{Sort}[\text{PValStudent}] - \frac{1}{365}\text{Range}[365]\right\}\right]\right]$$

```
PValKolmoSmirnov = KolmoSmirnov /@ Transpose[Transpose /@ SolarSplit[tempTX6IV, YearsIV] ];
ListPlot[PValKolmoSmirnov, Joined → True]
Mean[PValKolmoSmirnov]
```

Calculations on 11 - day filtered data

Calculations on 21 - day filtered data

- **21-day filter**

```
tempTN2 = MovingAverage[tempTN, 21];
tempTX2 = MovingAverage[tempTX, 21];
time2 = Drop[Drop[time, 10], -10];
```

■ Calculating and removing annual cycle (on 21-day filtered data)

- **First get rid of the 29 February to get 365-day years**

In order to get sets of years with equal length, the 29 February's are removed. This is perfomed using the DeleleCases function.

```
TT = Evaluate[DeleteCases[Evaluate[Transpose[{tempTN2, time2}]], {_Real, {_Integer, 2, 29}}]];
{tempTN7, time7} = Transpose[TT];
TT =
  Evaluate[DeleteCases[Evaluate[Transpose[{tempTX2, time2}]], {_Real, {_Integer, 2, 29}}]];
{tempTX7, time7} = Transpose[TT];
```

- **Remove incomplete last and first years**

1775 has 10 days missing after filtering and 2005 is incomplete. We remove these two years to keep a dataset of complete years.

```
tempTN8 = Take[Drop[tempTN7, 355], 229 * 365];
tempTX8 = Take[Drop[tempTX7, 355], 229 * 365];
time8 = Take[Drop[time7, 355], 229 * 365];
{time8[[1]], time8[[-1]]}

{{1776, 1, 1}, {2004, 12, 31}}
```

This last line shows the range of data from first to last date

- **Calculate and remove annual cycle**

It is not necessary to remove annual cycle in this study but it is useful to remove it to identify better the data discontinuities. This has no effect on the analysis that follow.

```
tempannual = Map[Mean, Transpose[Partition[tempTN8, 365]]];
TMP = Take[Nest[Join[#, #] &, tempannual, 8], 229 * 365];
tempTN9 = tempTN8 - TMP;
tempannual = Map[Mean, Transpose[Partition[tempTX8, 365]]];
TMP = Take[Nest[Join[#, #] &, tempannual, 8], 229 * 365];
tempTX9 = tempTX8 - TMP;
```

■ Selecting high and low solar years for 21-day data

We select here the High and Low ensembles according to LeMouel et al.

```
SolarCyclesF = {{3, 1776, 1784}, {4, 1785, 1798}, {5, 1799, 1810}, {6, 1811, 1823},
   {7, 1824, 1833}, {8, 1834, 1843}, {9, 1844, 1856}, {10, 1857, 1867}, {11, 1868, 1878},
   {12, 1879, 1889}, {13, 1890, 1901}, {14, 1902, 1913}, {15, 1914, 1923},
   {16, 1924, 1933}, {17, 1934, 1944}, {18, 1945, 1954}, {19, 1955, 1964},
   {20, 1965, 1976}, {21, 1977, 1986}, {22, 1987, 1996}, {23, 1997, 2004}};
HighF = {3, 4, 8, 9, 11, 18, 19, 20, 21, 22, 23};
HighYearsF = {{0, 1798 - 1776 + 1}, {1834 - 1776, 1856 - 1834 + 1},
   {1868 - 1776, 1878 - 1868 + 1}, {1945 - 1776, 2004 - 1945 + 1}};
LowF = {5, 6, 7, 10, 12, 13, 14, 15, 16, 17};
LowYearsF =
  {{1799 - 1776, 1833 - 1799 + 1}, {1857 - 1776, 1867 - 1857 + 1}, {1879 - 1776, 1944 - 1879 + 1}};
YearsF = {HighYearsF, LowYearsF}
```

```
{{{0, 23}, {58, 23}, {92, 11}, {169, 60}}, {{23, 35}, {81, 11}, {103, 66}}}
```

▪ Calculate the composite difference between high and low solar years for 21-day data

▪ Analysis of the Praha data

First we reconstruct the Fig .4 a of LMKC, in particular their estimate of the error. The reconstruction is based on an exchange with LeMouel. It was asked to him by E - mail how σ was evaluated in Fig .4 a. The response was received by (standard) mail and contained a calculation of the daily variance over the H and L subsets of days that is (using LeMouel notations and words)

$$\sigma_H{}^2 = \frac{1}{N_H}\sum_i (T_{iH} - T_H)^2$$ where T_{iH} is the temperature of the calendar day i,

T_H is the average of T_{iH},
N_H is the number of days entering the 21 - day averages for the H ensemble,
that is $N_H = 21 \times 119 = 2499$ for Praha. Similar expression holds for the L ensemble. This, obviously, is not the right answer since this quantity is a daily variance, which is of the order of 25 K^2 during winter, and not a vaiance for the mean. It, however, indicates that LMKC has used this quantity and our hypothesis is that they have estimated the variance of the average over H for each day by dividing this σ_H by N_H. This is what we calculate now (notice : N_H here not the same as N^H in our paper, in the paper $N^H = 119$)

Algorithm :
 - varTx2 : variance of the daily fluctuations over 21 - day intervals, annual cycle not removed
 - varTx7: filtering out of the 29 Febs from varTx2
 - varTx8: selection of the sequence of full years (1776-2004)
 - tx, vx: Splitting of the temperatures and daily variance into H and L ensembles , the result of SolarSplit is a three-level list, level 1 separates H and L ensembles, level 2 is the list of years, level 3 is a list of 365 values for each year
- Stdx: variance of the difference H-L, according to LMKC, as the sum of the two variances, each being calculated as the daily variance averaged over the ensemble divided by the number of days in the ensemble. In this calculation Variance and Mean are performed on the second level of the list for each day.

The annual cycle is not removed since there is no indication that this has been done in LMKC. This has no influence on the calculation of the variance of average temperatures tx and a weak influence on vx since 21 days is short with respect to the length of the year.

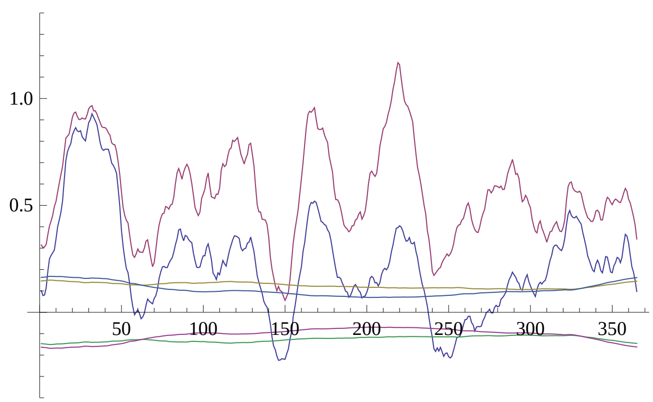
```
varTN2 = MovingAverage[tempTN^2, 21] - MovingAverage[tempTN, 21]^2;
varTX2 = MovingAverage[tempTX^2, 21] - MovingAverage[tempTX, 21]^2;
TT =
  Evaluate[DeleteCases[Evaluate[Transpose[{varTN2, time2}]], {_Real, {_Integer, 2, 29}}]];
{varTN7, dumb} = Transpose[TT];
TT =
  Evaluate[DeleteCases[Evaluate[Transpose[{varTX2, time2}]], {_Real, {_Integer, 2, 29}}]];
{varTX7, dunm} = Transpose[TT];
varTN8 = Take[Drop[varTN7, 355], 229 * 365];
varTX8 = Take[Drop[varTX7, 355], 229 * 365];
vN = SolarSplit[varTN8, YearsF];
vX = SolarSplit[varTX8, YearsF];
tN = SolarSplit[tempTN8, YearsF];
tX = SolarSplit[tempTX8, YearsF];
```

$$\text{StdN} = \sqrt{\frac{\text{Variance}[\text{tN}[[1]]] + \text{Mean}[\text{vN}[[1]]]}{21 * \text{Length}[\text{tN}[[1]]]} + \frac{\text{Variance}[\text{tN}[[2]]] + \text{Mean}[\text{vN}[[2]]]}{21 * \text{Length}[\text{tN}[[2]]]}};$$

$$\text{StdX} = \sqrt{\frac{\text{Variance}[\text{tX}[[1]]] + \text{Mean}[\text{vX}[[1]]]}{21 * \text{Length}[\text{tX}[[1]]]} + \frac{\text{Variance}[\text{tX}[[2]]] + \text{Mean}[\text{vX}[[2]]]}{21 * \text{Length}[\text{tX}[[2]]]}};$$

Fig .4 a of LMKC is then reconstructed using the H - L differences calculated for TN and TX and plotting ± Stdx

```
ListPlot[{SolarDiff[tempTN9, YearsF], SolarDiff[tempTX9, YearsF], StdX, -StdX, StdN, -StdN},
 Joined → True, PlotRange → {Automatic, {-0.4, 1.4}}]
```



This figure is plotted with the same scale as LMKC. The curves for TN and TX are visually identical to those shown in Fig.4a of LMKC
We believe our calculation is correct. Differences are in principle only in the first and last years which are removed in our calculations .
The "error" curves are also visually identical to those of LMKC, hence validating our hypothesis.

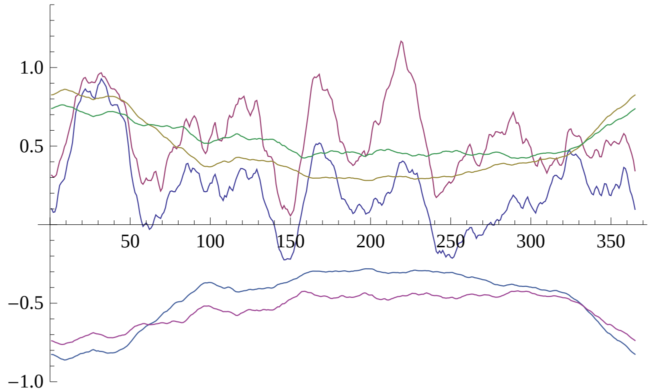
This is wrong for two reasons :
-1) because it cannot be assumed that the daily fluctuations of temperature are independent over each 21 - day sample, while it is legimate to make this hypothesis over two separate years; this leads to a large underestimate of σ, by a factor about 3 as we shall see below,
- 2) because ± σ is only a 68 % confidence interval for a Gaussian distribution, and it is necessary to plot ± 1.65 σ to get a standard 90 % confidence interval; in this case, it is still expected that about 10 % of the data can be outside of this interval without any statistical meaning
Hence, on the overall, confidence interval has been understimated by a factor near to 5 as checked in the next figure.

Calculation of the a factor multiplying the Student variance to get 90 % confidence interval

```
a = x /. FindRoot[NIntegrate[
    PDF[StudentTDistribution[Length[tempTN9] / 365 - 2], y], {y, 0, x}] == 0.45, {x, 1.6}]
```

```
1.65159
```

```
ListPlot[{SolarDiff[tempTN9, YearsF], SolarDiff[tempTX9, YearsF], a StdDev[tempTN9, YearsF],
  a StdDev[tempTX9, YearsF], -a StdDev[tempTN9, YearsF], -a StdDev[tempTX9, YearsF]},
 Joined → True, PlotRange → {Automatic, {-1, 1.4}}]
```



This is Fig.4a of LMKC as it should have been. Here we have calculated the bounds by calculating first the variance for the two subseries and summing them (approxim if two subsets of same size) and multiplying by 1.65 to get the 90 % confidence interval. The positive bias for the TX series is due to the contribution of the global warming as shown in nect subsection using an other, non-parametric, estimate of the confidence interval.
It has been checked that this result is totally independent of the removal of the annual cycle.
Due to the number of variables in teh average, the Student distribution could be replaced without any visually perceptible difference by the Gaussian limit distribution.

- **Figure for publication**

Calculate composite difference after bootstrapping over a random permutation of the years

Example of solar shift after resampling

```
ListPlot[SolarDiff[Shuffle[tempTX9, 229], YearsF], Joined → True]
```

We see that any permutaion of the years produces a composite difference with similar amplitude

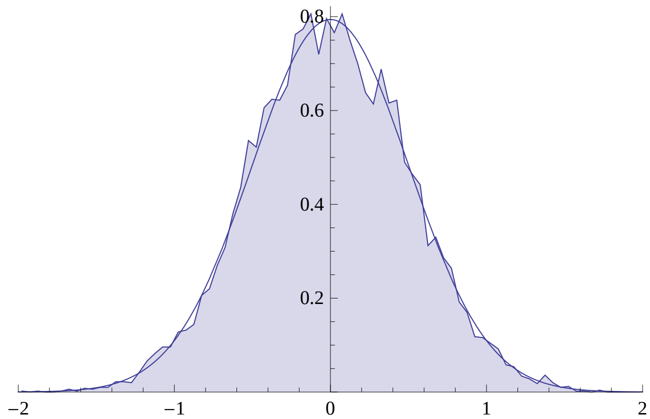
- **Generate a table of composite differences with 10000 elements**

Solar shift is calculated over 10000 resamplings of the dataset, the pdf is estimated by simple bining.

```
RefTN = Transpose[Table[SolarDiff[Shuffle[tempTN9, 229], YearsF], {i, 1, 10000}]];
RefTX = Transpose[Table[SolarDiff[Shuffle[tempTX9, 229], YearsF], {i, 1, 10000}]];
RefpdfTN = Map[Pdf, RefTN];
RefpdfTX = Map[Pdf, RefTX];
```

This is an example of the estimated pdf showing fairly nice Gaussian profile and that more refined calculation with larger number of drawings will probably not make a big change. The Gaussian is plotted with 0 mean and same variance as the estimated distribution.

```
std = √(Total[RefpdfTN[[2]] BinVal^2] BinSize) ;
g1 = ListPlot[Transpose[{BinVal, RefpdfTN[[2]]}],
   Joined → True, PlotRange → {{-2, 2}, {0, Automatic}}, Filling → 0];
g2 = Plot[PDF[NormalDistribution[0, std], x], {x, -2, 2}];
Show[g1, g2]
```

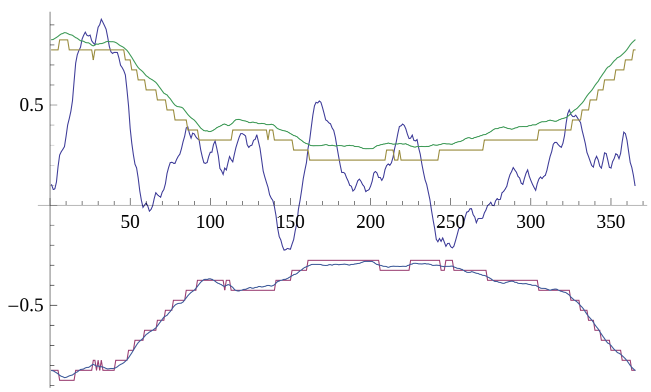


Let us now estimate the (5%-95%) two-sided confidence interval

```
Sig10TN = Confidence[RefpdfTN, 0.05];
Sig10TX = Confidence[RefpdfTX, 0.05];
```

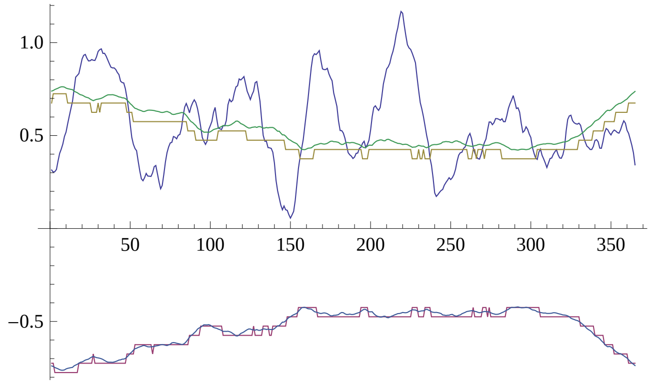
Plot, for TN, of the confidence interval estimated using both methods (resampling and from the Student distribution)

```
ListPlot[{SolarDiff[tempTN9, YearsF], Sig10TN[[1]], Sig10TN[[2]],
  a StdDev[tempTN9, YearsF], -a StdDev[tempTN9, YearsF]}, Joined → True]
```



Plot, for TX, of the confidence interval estimated using both methods (resampling and from the Student distribution)

```
ListPlot[{SolarDiff[tempTX9, YearsF], Sig10TX[[1]], Sig10TX[[2]],
  a StdDev[tempTX9, YearsF], -a StdDev[tempTX9, YearsF]}, Joined → True]
```



This shows that bootstrapping and the Student-t test provide almost excatly the same confidence interval.
This shows that the result of LeMouel et al on TN data is not significant even with all the cycles retained. The TX curve is outside the 90% interval more than 10% of its span, hence the positive value of the solar shift is in this case significant. However, we have the alternate hypothesis that this enterely due to the contribution of anthropogenic forcing during teh last 50 years of the data. We see below that indeed this hypothesis is not rejected by the data.

- **Figures for publication**

■ Analysis after removal of the last 3 cycles (ensemble III according to LeMouel et al.)

■ Analysis after removal of the last 5 cycles (ensemble IV according to LeMouel et al.)

We drop the last 5 cycles or 50 years corresponding to the period where anthropogenic forcing is felt

```
tempTN9IV = Drop[tempTN9, -50 * 365];
tempTX9IV = Drop[tempTX9, -50 * 365];
HighIV = {3, 4, 8, 9, 11, 18};
HighYearsIV = {{0, 1798 - 1776 + 1}, {1834 - 1776, 1856 - 1834 + 1},
   {1868 - 1776, 1878 - 1868 + 1}, {1945 - 1776, 1954 - 1945 + 1}};
YearsIV = {HighYearsIV, LowYearsF}
RefTNIV = Transpose[Table[SolarDiff[Shuffle[tempTN9IV, 179], YearsIV], {i, 1, 10000}]];
RefTXIV = Transpose[Table[SolarDiff[Shuffle[tempTX9IV, 179], YearsIV], {i, 1, 10000}]];
RefpdfTNIV = Map[Pdf, RefTNIV];
RefpdfTXIV = Map[Pdf, RefTXIV];
Sig10TNIV = Confidence[RefpdfTNIV, 0.05];
Sig10TXIV = Confidence[RefpdfTXIV, 0.05];
```

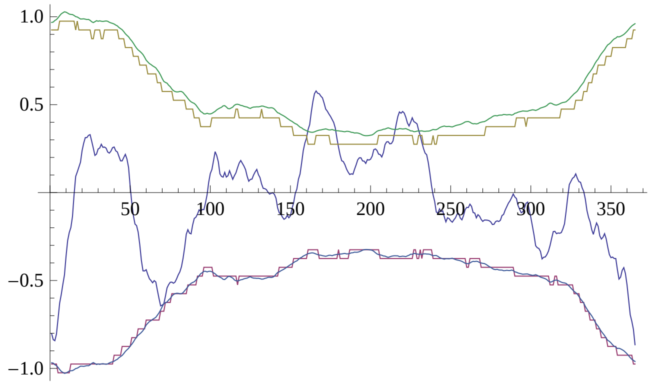
```
{{{0, 23}, {58, 23}, {92, 11}, {169, 10}}, {{23, 35}, {81, 11}, {103, 66}}}
```

```
aIV = x /. FindRoot[NIntegrate[
     PDF[StudentTDistribution[Length[tempTN9IV] / 365 - 2], y], {y, 0, x}] == 0.45, {x, 1.6}]
```

```
1.65351
```

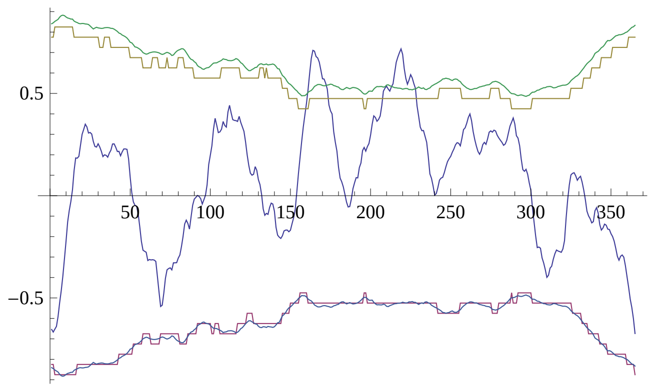
- **Solar shift for TN**

```
ListPlot[{SolarDiff[tempTN9IV, YearsIV], Sig10TNIV[[1]], Sig10TNIV[[2]],
  aIV StdDev[tempTN9IV, YearsIV], -aIV StdDev[tempTN9IV, YearsIV]}, Joined → True]
```



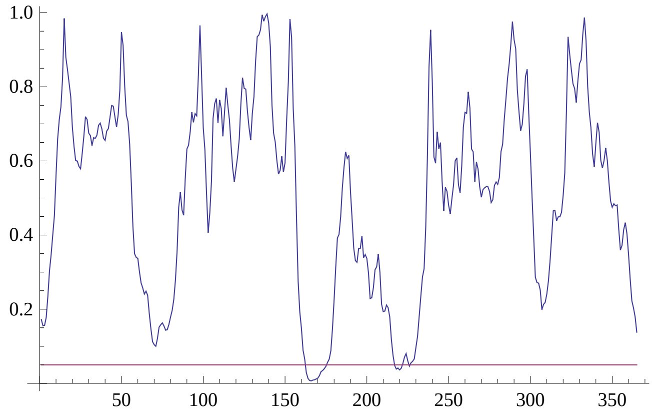
- **Solar shift for TX**

```
ListPlot[{SolarDiff[tempTX9IV, YearsIV], Sig10TXIV[[1]], Sig10TXIV[[2]],
  aIV StdDev[tempTX9IV, YearsIV], -aIV StdDev[tempTX9, YearsIV]}, Joined → True]
```



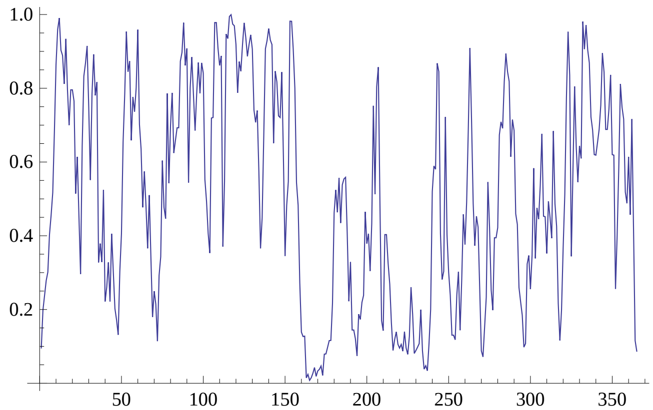
- **Student and Kolmogorov-Smirnov P-value for TN**

```
PValStudentTN = StudentTest /@ Transpose[Transpose /@ SolarSplit[tempTN9IV, YearsIV] ];
ListPlot[{PValStudentTN, ConstantArray[0.05, 365]}, Joined → True]
Mean[PValStudentTN]
```




```
0.508964
```

```
PValKolmoSmirnov = KolmoSmirnov /@ Transpose[Transpose /@ SolarSplit[tempTN9IV, YearsIV] ];
ListPlot[PValKolmoSmirnov, Joined → True]
Mean[PValKolmoSmirnov]
```



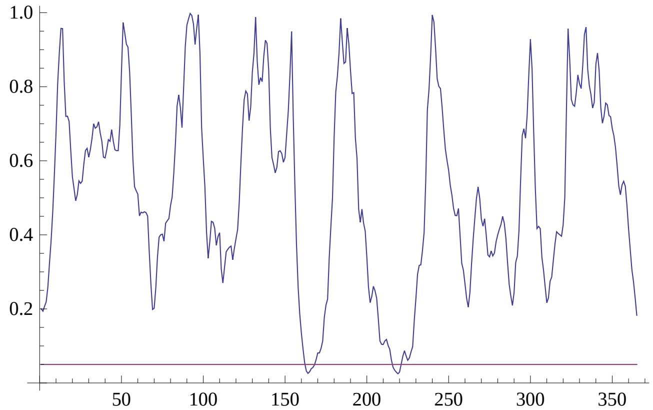

```
0.513825
```

The distribution of p - value is similar to a Brownian bridge

$$\texttt{ListPlot}\left[\texttt{Transpose}\left[\left\{\frac{1}{365}\texttt{Range[365]},\ \texttt{Sort[PValStudent]} - \frac{1}{365}\texttt{Range[365]}\right\}\right]\right]$$

- **Student and Kolmogorov-Smirnov P-value for TX**

```
PValStudentTX = StudentTest /@ Transpose[Transpose /@ SolarSplit[tempTX9IV, YearsIV] ];
ListPlot[{PValStudentTX, ConstantArray[0.05, 365]}, Joined → True]
Mean[PValStudentTX]
```




```
0.521503
```

```
PValKolmoSmirnov = KolmoSmirnov /@ Transpose[Transpose /@ SolarSplit[tempTX9IV, YearsIV] ];
Mean[PValKolmoSmirnov]
```

```
0.518891
```

Test that the assumption of equal variance has no impact

```
DiffPValStudentTX =
  (StudentTest2 /@ Transpose[Transpose /@ SolarSplit[tempTX9IV, YearsIV] ]) -
   (StudentTest /@ Transpose[Transpose /@ SolarSplit[tempTX9IV, YearsIV] ]);
ListPlot[DiffPValStudentTX, Joined → True]
Mean[DiffPValStudentTX]
```

ListPlot$\left[\text{Transpose}\left[\left\{\frac{1}{365}\text{Range}[365], \text{Sort}[\text{PValStudent}] - \frac{1}{365}\text{Range}[365]\right\}\right]\right]$

We see that the Kolmogorov - Smirnov test provides similar results to teh Student test but very noisy p - value. This test is in fact poorly adpated here since we only want to test that the two means are significantly different and since the samples contain 100 elements only, which is much too small to sample a distribution.

- **Figures for publication**

Calculations on 41 - day filtered data

Estimate of the oversampling effect by LMKC

- **Estimate of the oversampling factor**

```
Variance[MovingAverage[Shuffle[tempTN6, 230], 21]]
```

```
5.16441
```

- **I Resampling of the years done for each date over the whole dataset**

```
DailyShuffle = Flatten[Transpose[Map[RandomSample, Transpose[Partition[#, 365]]]]] &;
```

```
Variance[tempTN9]
```

```
5.36643
```

```
Variance[MovingAverage[tempTN6, 21]]
```

```
5.35785
```

```
Variance[tempTN9] / Variance[MovingAverage[DailyShuffle[tempTN6], 21]]
```

```
7.95589
```

We see that the variance of the 21 - day averages is reduced by a factor 8 when the years are resampled independently for each date before averaging

- **II Resampling of the years within the H and L ensembles, leaving the solar shift unchanged**

```
DailyShuffle2 = Flatten[Transpose[Map[RandomSample, Transpose[#]]]] &;
```

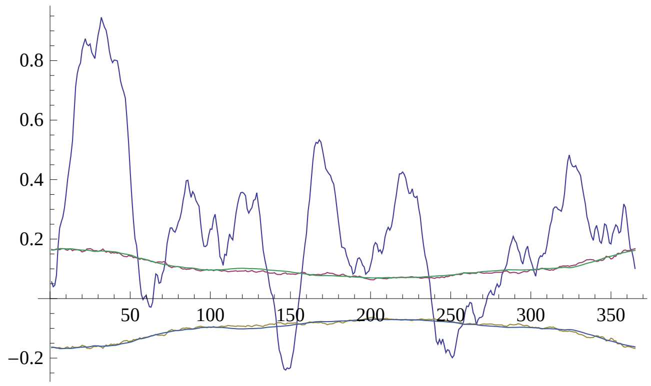
Here we start from the unfiltered data, we split them into the H and L ensemble (in AA),
we shuffle separately the years for each date within the H and L ensembles, and we pass the 21 - day filter on the result (AA1)
Then remove first and last incomplte years and partition into years and make two separate sets for H (AA2H) and L (AA2H) ensembles. The solar shift and its variance are calculated using these quantities and same formula as for the standard calculation.
This procedure leaves the solar shift unchanged up to side effects on the boundary of years and remval of 1 year with respect to data in tempTN9.
Running following cell requires to run preliminary calculations for unfiltered data, to calculate tempTN6 et tempTX6

```
AA = SolarSplit[tempTN6, Years];
{Length[AA[[1]]], Length[AA[[2]]]}
AA1 = MovingAverage[Flatten[{DailyShuffle2[AA[[1]]], DailyShuffle2[AA[[2]]]}], 21];
AA2 = Partition[Take[Drop[AA1, 355], 228 * 365], 365];
AA2HN = Take[AA2, 117]; AA2LN = Drop[AA2, 117];
SDAAN = Mean[AA2HN] - Mean[AA2LN];
```

$$\text{StdAAN} = \sqrt{\frac{116\,\text{Variance[AA2HN]} + 111\,\text{Variance[AA2LN]}}{226}\left(\frac{1}{118} + \frac{1}{111}\right)};$$

```
ListPlot[{SDAAN, StdAAN, -StdAAN, StdN, -StdN},
 Joined → True, PlotRange → All, PlotRange → {Automatic, {-0.4, 1.4}}]
Mean[StdDev[tempTN9, YearsF] / StdAAN]
```

{118, 112}



2.68182

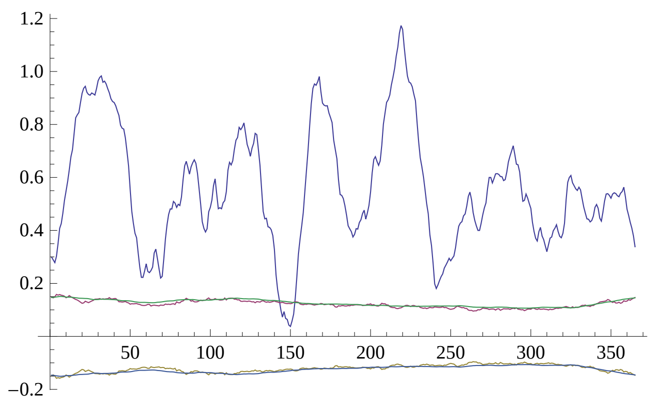
We see that the solar shift is preserved and that the sigma of LMKC is recovered due to the reduction of the variance of 21 - day averages when data are resampled within H and L ensembles. C.Q.F.D. This data to be added to the figures.

```
AA = SolarSplit[tempTX6, Years];
{Length[AA[[1]]], Length[AA[[2]]]}
AA1 = MovingAverage[Flatten[{DailyShuffle2[AA[[1]]], DailyShuffle2[AA[[2]]]}], 21];
AA2 = Partition[Take[Drop[AA1, 355], 228 * 365], 365];
AA2HX = Take[AA2, 117]; AA2LX = Drop[AA2, 117];
SDAAX = Mean[AA2HX] - Mean[AA2LX];
```

$$\mathtt{StdAAX} = \sqrt{\frac{116\,\mathtt{Variance[AA2HX]} + 111\,\mathtt{Variance[AA2LX]}}{226}\left(\frac{1}{118} + \frac{1}{111}\right)};$$

```
ListPlot[{SDAAX, StdAAX, -StdAAX, StdX, -StdX},
 Joined → True, PlotRange → All, PlotRange → {Automatic, {-0.4, 1.4}}]
Mean[StdDev[tempTX9, YearsF] / StdAAX]
```

```
{118, 112}
```




```
2.67442
```

This demonstatres that the sigma of LMKC is recovered using the correct formula after resampling in a way which preserves the solar shift but decorralates the daily temperatures by a permutation over the years for each date within each sub ensemble H and L. In other word, the solar shift found by LMKC would be significant and their calculation would be correct on an hyothetical planet with such a data series but not on the Earth.

Remark 1 : we do not do here any bootstrapping, a single permutaion is enough for the demonstration.

Remark 2 : if the permutation was done not within the ensembles H and L separately but over the whole data set, the solar shift would be killed or at least reduced to values unsignificant with respect to teh vnew confidence interval.

Estimate the autocorrelation of the daily temperature fluctuations

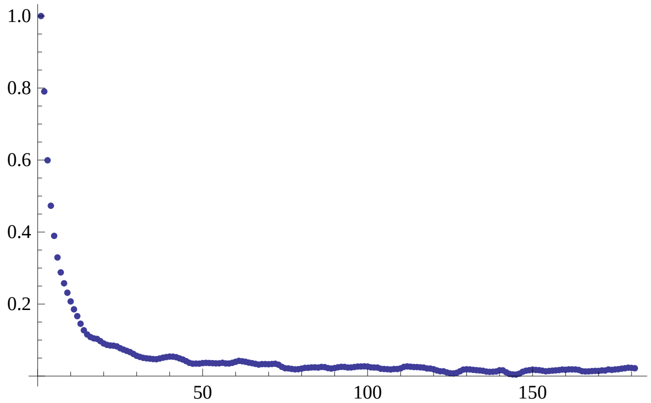
This is a crude procedure which is only meant to provide an order of magnitude of the correlation time
One might also be tempted to take the life time of LM (but we know it as a bad estimator for a non pure AR(1) process)

The resampling approach in previous subsection provides a much better estimate.

Remove last 50 years to eliminate trailing correlation due to climate shift (very limited effect anyway).

```
autocorrTN = Table[Correlation[
    Drop[Drop[tempTN6, 50 * 365], i], Drop[Drop[tempTN6, 50 * 365], -i]], {i, 0, 180}];
```

```
ListPlot[autocorrTN, PlotRange → All]
```




```
IntScaleTN =
 Integrate[Interpolation[Transpose[{Range[181] - 1, autocorrTN}], InterpolationOrder → 4][x],
  {x, 0, 90}]
```

```
8.17591
```

```
IntScaleTN =
 Integrate[Interpolation[Transpose[{Range[181] - 1, autocorrTN}], InterpolationOrder → 4][x],
  {x, 0, 150}]
```

```
9.2569
```

Oversampling factor

```
ListPlot[1.64 StdDev[tempTX9, YearsF] / StdX]
```

Using the series from Bologna since 1814

▪ Read data and preprocess them

▪ read data

TN : minimum temperature
TX : maximum temperature
TG: mean temperature
data are read directly from the file provided by ECAD after reformatting of data separators in order to accommodate ReadList requirements (coma replaced by white space)

```
snm = OpenRead[SrcDir <> "TN_nonblend_Bologna.datm"];
Skip[snm, String, 14]; (* skip header *)
data = Transpose[ReadList[snm, {Number, Number, Number, Number}]] /. -9999 → Missing[];
(* read data until the end of the file *)
Close[snm];
time = data[[2]];
tempTN = 0.1 * data[[3]] /. x___ Missing[] → Missing[] ;
snm = OpenRead[SrcDir <> "TX_nonblend_Bologna.datm"];
Skip[snm, String, 14]; (* skip header *)
data = Transpose[ReadList[snm, {Number, Number, Number, Number}]] /. -9999 → Missing[];
(* read data until the end of the file *)
Close[snm];
tempTX = 0.1 * data[[3]] /. x___ Missing[] → Missing[] ;
snm = OpenRead[SrcDir <> "TG_nonblend_Bologna.datm"];
Skip[snm, String, 14]; (* skip header *)
data = Transpose[ReadList[snm, {Number, Number, Number, Number}]] /. -9999 → Missing[];
(* read data until the end of the file *)
Close[snm];
tempTG = 0.1 * data[[3]] /. x___ Missing[] → Missing[] ;
year = Quotient[time, 10 000]; time = time - 10 000 year;
month = Quotient[time, 100]; day = time - 100 month;
time = Transpose[{year, month, day}];
```

- **Eliminate missing data during the last years**

```
tempTN = Take[tempTN, 69 396];
tempTX = Take[tempTX, 69 396];
tempTG = Take[tempTG, 69 396];
time = Take[time, 69 396];
{Length[Cases[tempTN, _Missing]],
 Length[Cases[tempTX, _Missing]], Length[Cases[tempTX, _Missing]]}
```

```
{2, 4, 4}
```

- **Fill missing data with interpolation from surrounding dates**

```
Cases[Transpose[{Range[69 396], tempTX}], {_Integer, _Missing}]
Cases[Transpose[{Range[69 396], tempTN}], {_Integer, _Missing}]
Cases[Transpose[{Range[69 396], tempTG}], {_Integer, _Missing}]
```

```
{{39 249, Missing[]}, {58 213, Missing[]}, {68 877, Missing[]}, {68 878, Missing[]}}

{{68 877, Missing[]}, {68 878, Missing[]}}

{{39 249, Missing[]}, {58 213, Missing[]}, {68 877, Missing[]}, {68 878, Missing[]}}
```

```
{time[[39 249]], time[[58 213]], time[[68 877]], time[[68 878]]}
```

```
{{1921, 6, 17}, {1973, 5, 19}, {2002, 7, 30}, {2002, 7, 31}}
```

```
tempTX[[39 249]] = 0.5 (tempTX[[39 248]] + tempTX[[39 250]]);
tempTX[[58 213]] = 0.5 (tempTX[[58 212]] + tempTX[[58 214]]);
tempTX[[68 877]] = (2 * tempTX[[68 876]] + tempTX[[68 879]]) / 3;
tempTX[[68 878]] = (tempTX[[68 876]] + 2 * tempTX[[68 879]]) / 3;
tempTG[[39 249]] = 0.5 (tempTG[[39 248]] + tempTG[[39 250]]);
tempTG[[58 213]] = 0.5 (tempTG[[58 212]] + tempTG[[58 214]]);
tempTG[[68 877]] = (2 * tempTG[[68 876]] + tempTG[[68 879]]) / 3;
tempTG[[68 878]] = (tempTG[[68 876]] + 2 * tempTG[[68 879]]) / 3;
tempTN[[68 877]] = (2 * tempTX[[68 876]] + tempTN[[68 879]]) / 3;
tempTN[[68 878]] = (tempTN[[68 876]] + 2 * tempTN[[68 879]]) / 3;
{Length[Cases[tempTN, _Missing]],
 Length[Cases[tempTX, _Missing]], Length[Cases[tempTG, _Missing]]}
```

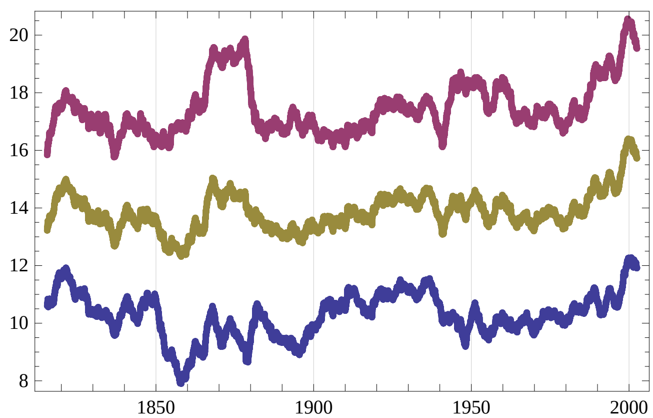
```
{0, 0, 0}
```

No data is marked as missing after removal of the last years and interpolation of the few gaps

- **3-year filter**

```
tempTN3 = MovingAverage[tempTN, 3 * 365];
tempTX3 = MovingAverage[tempTX, 3 * 365];
tempTG3 = MovingAverage[tempTG, 3 * 365];
time3 = Drop[Drop[time, 365 + 182], -365 - 182];
```

```
DateListPlot[
 {Transpose[{time3, tempTN3}], Transpose[{time3, tempTX3}], Transpose[{time3, tempTG3}]}]
```



This is the figure 1 of LeMouel et al
It is very strange that the maximum and minmum temperature are so badly correlated on this scale while they are at the daily scale

- **Calculating and removing annual cycle (on 21-day filtered data)**

- **21-day filter**

```
tempTN2 = MovingAverage[tempTN, 21];
tempTX2 = MovingAverage[tempTX, 21];
tempTG2 = MovingAverage[tempTG, 21];
time2 = Drop[Drop[time, 10], -10];
```

- **Get rid of the 29 February to make 365-day years**

```
TT = Evaluate[DeleteCases[Evaluate[Transpose[{tempTN2, time2}]], {_Real, {_Integer, 2, 29}}]];
{tempTN7, time7} = Transpose[TT];
TT =
  Evaluate[DeleteCases[Evaluate[Transpose[{tempTX2, time2}]], {_Real, {_Integer, 2, 29}}]];
{tempTX7, time7} = Transpose[TT];
TT =
  Evaluate[DeleteCases[Evaluate[Transpose[{tempTG2, time2}]], {_Real, {_Integer, 2, 29}}]];
{tempTG7, time7} = Transpose[TT];
```

- **Remove incomplete last and first years**

1804 has 10 days missing after filtering and 2003 is incomplete

```
tempTN8 = Take[Drop[tempTN7, 355], 188 * 365];
tempTX8 = Take[Drop[tempTX7, 355], 188 * 365];
tempTG8 = Take[Drop[tempTG7, 355], 188 * 365];
time8 = Take[Drop[time7, 355], 188 * 365];
{time8[[1]], time8[[-1]]}
```

```
{{1815, 1, 1}, {2002, 12, 31}}
```

This last line shows the range of data from first to last date

- **Calculate and remove annual cycle**

It is not necessary to remove annual cycle in this study but it is useful to remove it to identify better data discontinuities. This has no effect on the analysis that follow.

```
tempannual = Map[Mean, Transpose[Partition[tempTN8, 365]]];
TMP = Take[Nest[Join[#, #] &, tempannual, 8], 188 * 365];
tempTN9 = tempTN8 - TMP;
tempannual = Map[Mean, Transpose[Partition[tempTX8, 365]]];
TMP = Take[Nest[Join[#, #] &, tempannual, 8], 188 * 365];
tempTX9 = tempTX8 - TMP;
tempannual = Map[Mean, Transpose[Partition[tempTG8, 365]]];
TMP = Take[Nest[Join[#, #] &, tempannual, 8], 188 * 365];
tempTG9 = tempTG8 - TMP;
```

- **Looking at the bump**

- **Selecting high and low solar years for 21-day filtered data**

```
SolarCyclesF = {{6, 1815, 1823}, {7, 1824, 1833}, {8, 1834, 1843}, {9, 1844, 1856},
   {10, 1857, 1867}, {11, 1868, 1878}, {12, 1879, 1889}, {13, 1890, 1901}, {14, 1902, 1913},
   {15, 1914, 1923}, {16, 1924, 1933}, {17, 1934, 1944}, {18, 1945, 1954}, {19, 1955, 1964},
   {20, 1965, 1976}, {21, 1977, 1986}, {22, 1987, 1996}, {23, 1997, 2002}};
HighF = {8, 9, 11, 18, 19, 20, 21, 22, 23};
HighYearsF =
 {{1834 - 1815, 1856 - 1834 + 1}, {1868 - 1815, 1878 - 1868 + 1}, {1945 - 1815, 2002 - 1945 + 1}}
LowF = {6, 10, 12, 13, 14, 15, 16, 17};
LowYearsF = {{0, 1833 - 1815 + 1}, {1857 - 1815, 1867 - 1857 + 1}, {1879 - 1815, 1944 - 1879 + 1}}
YearsF = {HighYearsF, LowYearsF}
```

```
{{19, 23}, {53, 11}, {130, 58}}

{{0, 19}, {42, 11}, {64, 66}}

{{{19, 23}, {53, 11}, {130, 58}}, {{0, 19}, {42, 11}, {64, 66}}}
```

- **Calculate the composite difference between high and low solar years**

- **Variance of the filtered data**

■ Analysis of the solar shift

```
varTN2 = MovingAverage[tempTN^2, 21] - MovingAverage[tempTN, 21]^2;
varTX2 = MovingAverage[tempTX^2, 21] - MovingAverage[tempTX, 21]^2;
varTG2 = MovingAverage[tempTG^2, 21] - MovingAverage[tempTG, 21]^2;
TT =
  Evaluate[DeleteCases[Evaluate[Transpose[{varTN2, time2}]], {_Real, {_Integer, 2, 29}}]];
{varTN7, dumb} = Transpose[TT];
TT =
  Evaluate[DeleteCases[Evaluate[Transpose[{varTX2, time2}]], {_Real, {_Integer, 2, 29}}]];
{varTX7, dunm} = Transpose[TT];
TT =
  Evaluate[DeleteCases[Evaluate[Transpose[{varTG2, time2}]], {_Real, {_Integer, 2, 29}}]];
{varTG7, dunm} = Transpose[TT];
varTN8 = Take[Drop[varTN7, 355], 188 * 365];
varTX8 = Take[Drop[varTX7, 355], 188 * 365];
varTG8 = Take[Drop[varTG7, 355], 188 * 365];
vN = SolarSplit[varTN8, YearsF];
vX = SolarSplit[varTX8, YearsF];
vG = SolarSplit[varTG8, YearsF];
tN = SolarSplit[tempTN8, YearsF];
tX = SolarSplit[tempTX8, YearsF];
tG = SolarSplit[tempTG8, YearsF];
```

$$\text{StdN} = \sqrt{\frac{\text{Variance[tN[[1]]]} + \text{Mean[vN[[1]]]}}{21 * \text{Length[tN[[1]]]}} + \frac{\text{Variance[tN[[2]]]} + \text{Mean[vN[[2]]]}}{21 * \text{Length[tN[[2]]]}}};$$

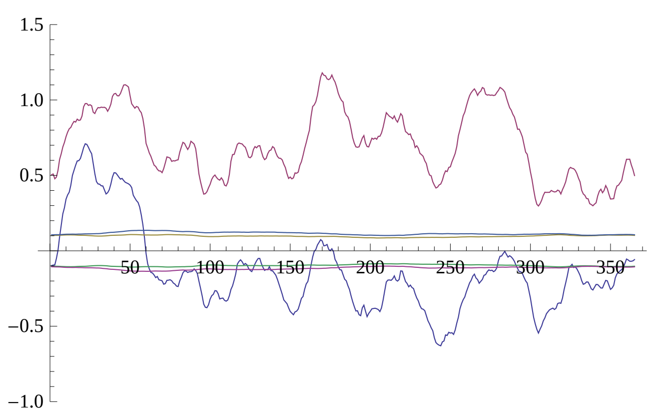
$$\text{StdX} = \sqrt{\frac{\text{Variance[tX[[1]]]} + \text{Mean[vX[[1]]]}}{21 * \text{Length[tX[[1]]]}} + \frac{\text{Variance[tX[[2]]]} + \text{Mean[vX[[2]]]}}{21 * \text{Length[tX[[2]]]}}};$$

$$\text{StdG} = \sqrt{\frac{\text{Variance[tG[[1]]]} + \text{Mean[vG[[1]]]}}{21 * \text{Length[tX[[1]]]}} + \frac{\text{Variance[tG[[2]]]} + \text{Mean[vG[[2]]]}}{21 * \text{Length[tX[[2]]]}}};$$

```
a =
 x /. FindRoot[NIntegrate[PDF[StudentTDistribution[Length[tX[[1]]] + Length[tX[[2]]] - 2], y],
     {y, 0, x}] == 0.45, {x, 1.6}]

1.65309
```

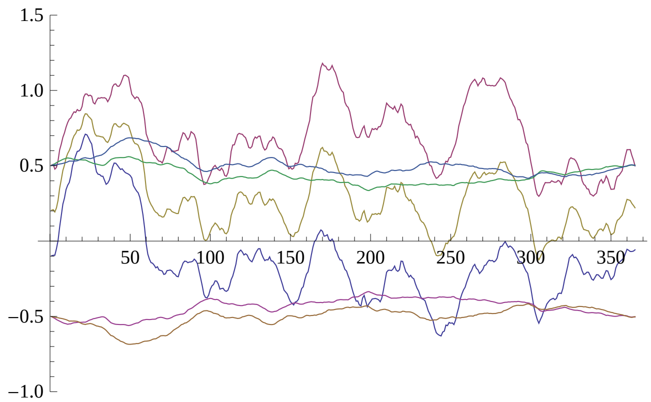
This is figure 4b of LeMouel et al.

```
ListPlot[{SolarDiff[tempTN9, YearsF], SolarDiff[tempTX9, YearsF], StdN, -StdN, StdX, -StdX},
 Joined → True, PlotRange → {Automatic, {-1, 1.5}}]
```



This is extremely close to fig4b of LeMouel et al. Our calculation does not use two years at the beginnig and the end of the dataset. Close inspection reveals that their error bar is perhaps slightly smaller.

```
ListPlot[{SolarDiff[tempTN9, YearsF], SolarDiff[tempTX9, YearsF], SolarDiff[tempTG9, YearsF],
  a StdDev[tempTN9, YearsF], a StdDev[tempTX9, YearsF], -a StdDev[tempTN9, YearsF],
  -a StdDev[tempTX9, YearsF]}, Joined → True, PlotRange → {Automatic, {-1, 1.5}}]
```



This is the figure as it should be. Here we have calculated the bounds by calculating first the variance for the two subseries and summing them (approxim if two subsets of same size) and multiplying by a to get the 90 % confidence interval. The positive bias for the TX series is due to the contribution of the global warming and of the 1867-1880 episode as shown in next subsection using an other, non-parametric, estimate of the confidence interval.

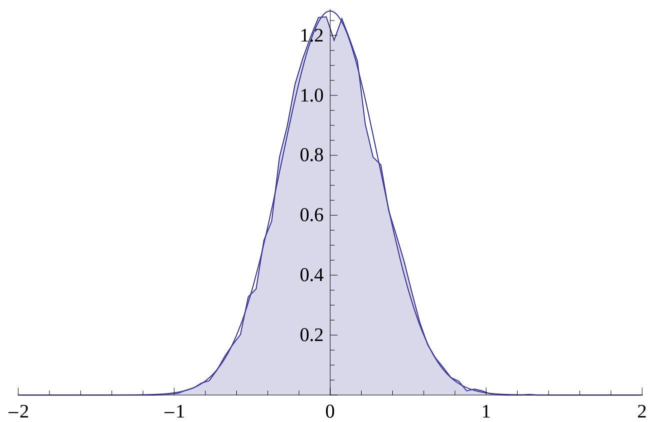
▪ Calculate composite difference after resampling randomly the years

▪ Generate a table of composite differences with 10000 elements

```
RefTN = Transpose[Table[SolarDiff[Shuffle[tempTN9, 188], YearsF], {i, 1, 10000}]];
RefTX = Transpose[Table[SolarDiff[Shuffle[tempTX9, 188], YearsF], {i, 1, 10000}]];
RefTG = Transpose[Table[SolarDiff[Shuffle[tempTG9, 188], YearsF], {i, 1, 10000}]];
RefpdfTN = Map[Pdf, RefTN];
RefpdfTX = Map[Pdf, RefTX];
RefpdfTG = Map[Pdf, RefTG];
```

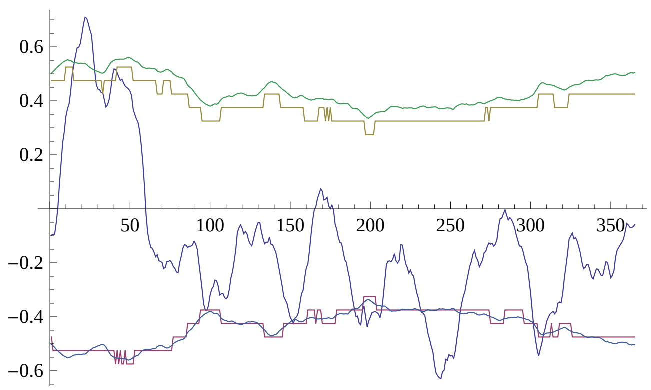
Check visually for Gaussianity on an example

```
std = Sqrt[Total[RefpdfTN[[2]] BinVal^2] BinSize];
g1 = ListPlot[Transpose[{BinVal, RefpdfTN[[2]]}],
   Joined → True, PlotRange → {{-2, 2}, {0, Automatic}}, Filling → 0];
g2 = Plot[PDF[NormalDistribution[0, std], x], {x, -2, 2}];
Show[g1, g2]
```

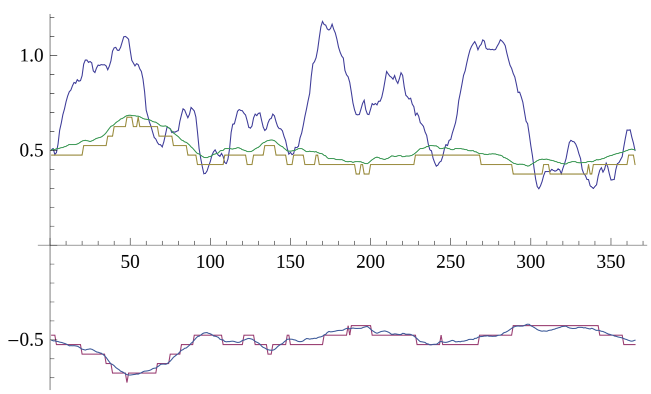


```
Sig10TN = Confidence[RefpdfTN, 0.05];
Sig10TX = Confidence[RefpdfTX, 0.05];
Sig10TG = Confidence[RefpdfTG, 0.05];
```

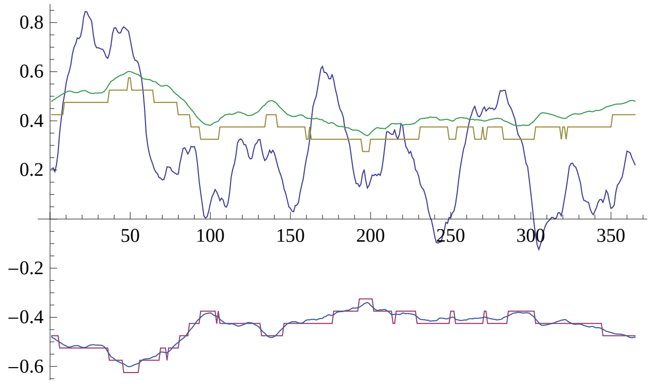
```
ListPlot[{SolarDiff[tempTN9, YearsF], Sig10TN[[1]], Sig10TN[[2]],
  a StdDev[tempTN9, YearsF], -a StdDev[tempTN9, YearsF]}, Joined → True]
```



```
ListPlot[{SolarDiff[tempTX9, YearsF], Sig10TX[[1]], Sig10TX[[2]],
  a StdDev[tempTX9, YearsF], -a StdDev[tempTX9, YearsF]}, Joined → True]
```



```
ListPlot[{SolarDiff[tempTG9, YearsF], Sig10TG[[1]], Sig10TG[[2]],
  a StdDev[tempTG9, YearsF], -a StdDev[tempTG9, YearsF]}, Joined → True]
```



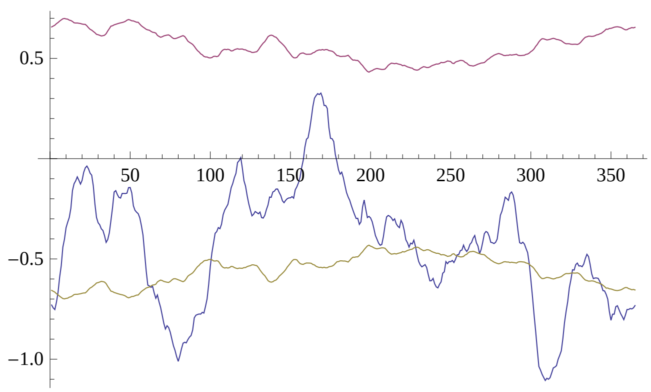
This shows that the result of LeMouel et al on TN and TG data is not significant. The TX curve is outside the significance interval

- **Analysis after removal of the last 3 cycles (ensemble III according to LeMouel et al.)**

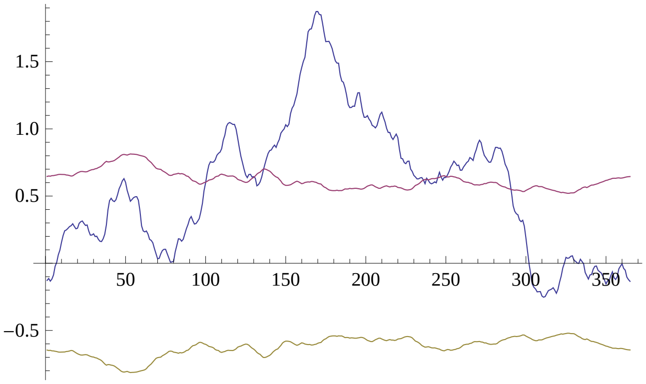
- **Analysis after removal of the last 5 cycles (ensemble IV according to LeMouel et al.)**

We drop the last 5 cycles or 48 years corresponding to the period where global warming is felt

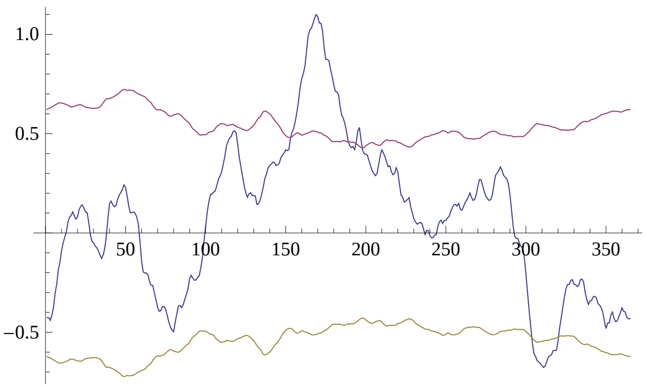
```
tempTN9IV = Drop[tempTN9, -48 * 365];
tempTX9IV = Drop[tempTX9, -48 * 365];
tempTG9IV = Drop[tempTG9, -48 * 365];

HighIV = {8, 9, 11, 18};
HighYearsIV =
 {{1834 - 1815, 1856 - 1834 + 1}, {1868 - 1815, 1878 - 1868 + 1}, {1945 - 1815, 1954 - 1945 + 1}}
YearsIV = {HighYearsIV, LowYearsF}

{{19, 23}, {53, 11}, {130, 10}}

{{{19, 23}, {53, 11}, {130, 10}}, {{0, 19}, {42, 11}, {64, 66}}}

aIV = x /. FindRoot[NIntegrate[
     PDF[StudentTDistribution[Length[tempTN9IV] / 365 - 2], y], {y, 0, x}] == 0.45, {x, 1.6}]

1.65597

ListPlot[{SolarDiff[tempTN9IV, YearsIV],
  aIV StdDev[tempTN9IV, YearsIV], -aIV StdDev[tempTN9IV, YearsIV]}, Joined → True]
```



```
ListPlot[{SolarDiff[tempTX9IV, YearsIV],
  aIV StdDev[tempTX9IV, YearsIV], -aIV StdDev[tempTX9IV, YearsIV]}, Joined → True]
```




```
ListPlot[{SolarDiff[tempTG9IV, YearsIV],
  aIV StdDev[tempTG9IV, YearsIV], -aIV StdDev[tempTG9IV, YearsIV]}, Joined → True]
```



We see that the TX series resists to the elimination of the last fifty years. This is due to the coincidence of the 1867 - 1880 episode with a high value of the solar cycle. To be demonstrated below by analysing the homogeneised series.

- **Analysis after removal of the last 5 cycles (ensemble IV according to LeMouel et al.) and of the XIX century bump (cycle 11)**

Processing the homogeneized series for Bologna

- **Reading the data**

```
snm = OpenRead[SrcDir <> "MEANMBOL.datm"];
data = Transpose[ReadList[snm, ConstantArray[Number, 13]] ];
 Close[snm];
years = data[[1]]; Nyears = Length[years]
tempH = Flatten[Transpose[Drop[data, 1]]];
timeH = Transpose[{Flatten[Transpose[Array[years &, 12]]],
    Flatten[Array[Range [12] &, Nyears]], ConstantArray[15, 12 * Nyears]}];
```

```
193
```

Comparison of homogeneized and non homogeneized series

```
timeH3y = Transpose[{Drop[Drop[Flatten[Transpose[Array[years &, 12]]], 18], -17], Drop[
      Drop[Flatten[Array[Range [12] &, Nyears]], 18], -17], ConstantArray[1, 12 * Nyears - 35]}];

DateListPlot[{Transpose[{time3, tempTG3}], Transpose[{timeH3y, MovingAverage[tempH, 36]}]}]
```

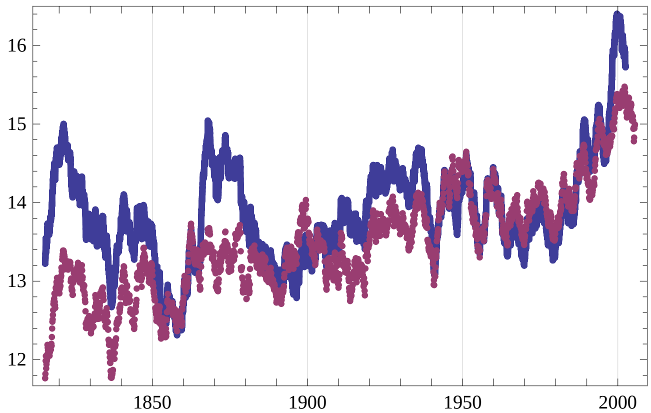


Figure for publication

Definition of the solar mask for the homogeneized record

```
SolarCyclesH = {{6, 1814, 1823}, {7, 1824, 1833}, {8, 1834, 1843}, {9, 1844, 1856},
    {10, 1857, 1867}, {11, 1868, 1878}, {12, 1879, 1889}, {13, 1890, 1901}, {14, 1902, 1913},
    {15, 1914, 1923}, {16, 1924, 1933}, {17, 1934, 1944}, {18, 1945, 1954}, {19, 1955, 1964},
    {20, 1965, 1976}, {21, 1977, 1986}, {22, 1987, 1996}, {23, 1997, 2006}};
HighH = {8, 9, 11, 18, 19, 20, 21, 22, 23};
HighYearsH =
  {{1834 - 1814, 1856 - 1834 + 1}, {1868 - 1814, 1878 - 1868 + 1}, {1945 - 1814, 2006 - 1945 + 1}}
LowH = {6, 10, 12, 13, 14, 15, 16, 17};
LowYearsH = {{0, 1833 - 1814 + 1}, {1857 - 1814, 1867 - 1857 + 1}, {1879 - 1814, 1944 - 1879 + 1}}
YearsH = {HighYearsH, LowYearsH}

{{20, 23}, {54, 11}, {131, 62}}

{{0, 20}, {43, 11}, {65, 66}}

{{{20, 23}, {54, 11}, {131, 62}}, {{0, 20}, {43, 11}, {65, 66}}}
```

■ Methods for monthly data

```
ShuffleM[dd_, Ns_] := Flatten[RandomSample[Partition[dd, 12], Ns]];
GatherYearM[temp_, {dl_, ll_}] := Take[Drop[temp, dl * 12], ll * 12]
SolarSplitM[temp_, years_] := Module[{HighTemp, LowTemp},
  HighTemp = Partition[Flatten[MapThread[GatherYearM[temp, #] &, {years[[1]]}]], 12];
  LowTemp = Partition[Flatten[MapThread[GatherYearM[temp, #] &, {years[[2]]}]], 12];
  Return[{HighTemp, LowTemp}]]
SolarDiffM[temp_, years_] := Mean[#[[1]]] - Mean[#[[2]]] & @@ {SolarSplitM[temp, years]}
StdDevM[temp_, years_] :=
```

$$\sqrt{\left(\frac{(\text{Length}[\#[[1]]]-1)\,\text{Variance}[\#[[1]]]+(\text{Length}[\#[[2]]]-1)\,\text{Variance}[\#[[2]]]}{\text{Length}[\#[[1]]]+\text{Length}[\#[[2]]]-2}\left(\frac{1}{\text{Length}[\#[[1]]]}+\frac{1}{\text{Length}[\#[[2]]]}\right)\right)}\ \&\ @@\ \{\text{SolarSplitM}[\text{temp},\ \text{years}]\}$$

■ Difference between high and low solar cycle

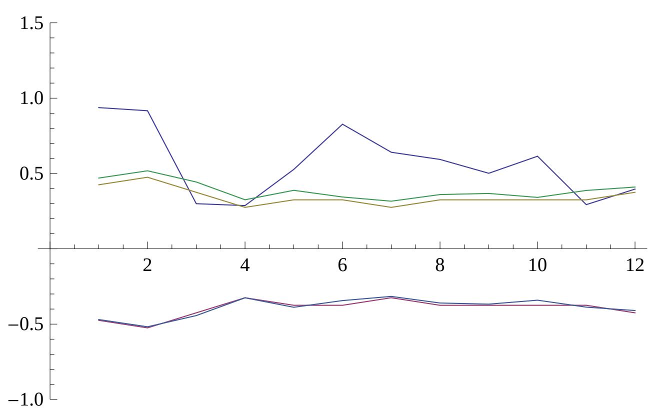
■ Using all the cycles

```
RefH = Transpose[Table[SolarDiffM[ShuffleM[tempH, Nyears], YearsH], {i, 1, 10 000}]];
RefpdfH = Map[Pdf, RefH];
Sig10H = Confidence[RefpdfH, 0.05];
```

```
a = x /. FindRoot[
   NIntegrate[PDF[StudentTDistribution[Length[tempH] / 12 - 2], y], {y, 0, x}] == 0.45, {x, 1.6}]
```

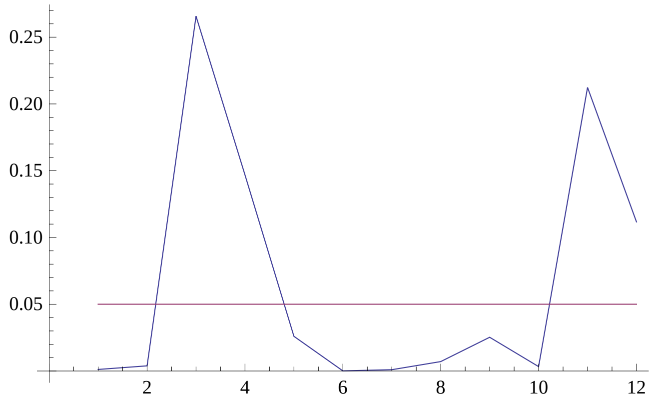
```
1.65287
```

```
ListPlot[{SolarDiffM[tempH, YearsH], Sig10H[[1]], Sig10H[[2]], a StdDevM[tempH, YearsH],
  -a StdDevM[tempH, YearsH]}, Joined → True, PlotRange → {Automatic, {-1, 1.5}}]
```



Student t - test

```
PValStudent = StudentTest /@ Transpose[Transpose /@ SolarSplitM[tempH, YearsH] ];
ListPlot[{PValStudent, ConstantArray[0.05, 12]}, Joined → True, PlotRange → All]
Mean[PValStudent]
```



```
0.0670208
```

- **Removing the last 5 cycles**

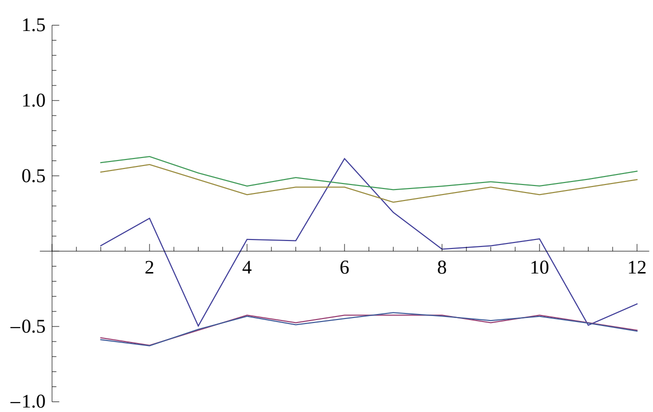
```
tempHIV = Drop[tempH, -12 * 52];

HighHIV = {8, 9, 11, 18}; HighYearsHIV =
 {{1834 - 1814, 1856 - 1834 + 1}, {1868 - 1814, 1878 - 1868 + 1}, {1945 - 1814, 1954 - 1945 + 1}}
YearsHIV = {HighYearsHIV, LowYearsH}
```

```
{{20, 23}, {54, 11}, {131, 10}}

{{{20, 23}, {54, 11}, {131, 10}}, {{0, 20}, {43, 11}, {65, 66}}}
```

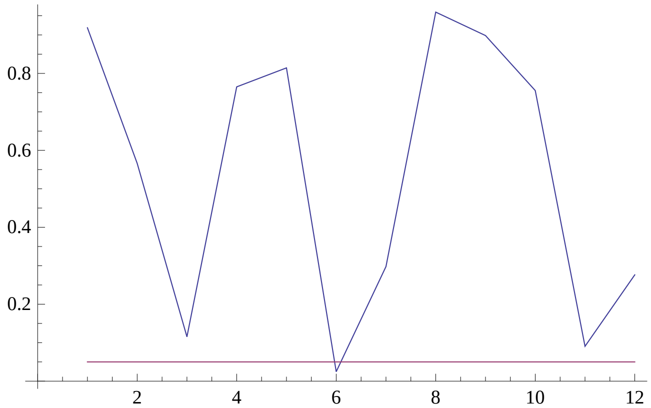
```
RefHIV =
  Transpose[Table[SolarDiffM[ShuffleM[tempHIV, Nyears - 52], YearsHIV], {i, 1, 10 000}]];
RefpdfHIV = Map[Pdf, RefHIV];
Sig10HIV = Confidence[RefpdfHIV, 0.05];

aIV = x /.
  FindRoot[NIntegrate[PDF[StudentTDistribution[Length[tempHIV] / 12 - 2], y], {y, 0, x}] == 0.45,
   {x, 1.6}]
```

```
1.65589
```

```
ListPlot[{SolarDiffM[tempHIV, YearsHIV], Sig10HIV[[1]], Sig10HIV[[2]],
  aIV StdDevM[tempHIV, YearsHIV], -aIV StdDevM[tempHIV, YearsHIV]},
 Joined → True, PlotRange → {Automatic, {-1, 1.5}}]
```



```
PValStudentIV = StudentTest /@ Transpose[Transpose /@ SolarSplitM[tempHIV, YearsHIV] ];
ListPlot[{PValStudentIV, ConstantArray[0.05, 12]}, Joined → True]
Mean[PValStudentIV]
```




```
0.540169
```

- **Figures for publication**