Subject: how to speed up multiple regressions?
Posted by Charlotte DeMott on Mon, 30 Apr 2001 17:18:10 GMT
View Forum Message <> Reply to Message

Hi,

I have some code to construct a composite of a meteorological phenomena in three dimensions (x, y, lag). The compositing index is a time series (ts) of a certain variable, and the data being composited (x, y, time) is regressed onto this compositing index. Because of the length of the time series and the size of the data array, and the fact that I do this compositing for multiple fields, I'm looking for ways to speed up the process, which is currently quite time consuming. The greatest amount of time seems to be spent in computing the significance of the correlation, rather than in computing the regressions. The regression is only done for periods where the signal is the "ts" time series is "big" (i.e., big = WHERE(ts GE threshold)).

Here are the main chunks of code used:

1) to do the regressions:

```
  ts_ac = A_CORRELATE(ts,lags) ; auto-corr of index time
series
  dataf = fltarr(dim(1),dim(2),2*lagdays+1,2) ; regression
a,b coefficients
  datar = fltarr(dim(1),dim(2),2*lagdays+1) ; corr. during
big var periods
  data_ac = fltarr(dim(1),dim(2),2*lagdays+1) ; data
auto_correlation
  data_tau = fltarr(dim(1),dim(2),2*lagdays+1) ;
decorrelation time scale
    ; Livezey & Chen, MWR '83
  print, 'computing regression coefficients...'
  FOR j = 0,dim(2)-1 DO BEGIN
    FOR i = 0,dim(1)-1 DO BEGIN
     temp = A_CORRELATE(data(i,j,*),lags)
     data_ac(i,j,*) = temp
     FOR lag = 0,2*lagdays DO BEGIN ; first = -29, last =
29
       dataf(i,j,lag,*) =
LINFIT(ts(big),data(i,j,big+lag-lagdays))
       datar(i,j,lag) =
CORRELATE(ts(big),data(i,j,big+lag-lagdays))
       data_tau(i,j,lag) = $
```

```
       (1.+2.*TOTAL(ts_ac(0:lag)*data_ac(i,j,0:lag))) > 1.
      ENDFOR
     ENDFOR
   ENDFOR
```

2) to compute the significance of the correlation:

```
   ; compute the number of degress of freedom
   datadof =
(fltarr(dim(1),dim(2),2*lagdays+1)+big_count)/data_tau
   ; find where corrlation is significant at 95% level
(Student's t)
  data_sig = intarr(dim(1),dim(2),2*lagdays+1)
   data_t = fltarr(dim(1),dim(2),2*lagdays+1)
   tsval = 2.*SQRT(mean_var)
   datacomp = fltarr(dim(1),dim(2),2*lagdays+1)
   FOR lag = 0,2*lagdays DO BEGIN
     FOR j = 0,dim(2)-1 DO BEGIN
      FOR i = 0,dim(1)-1 DO BEGIN
        data_t(i,j,lag) =
((ABS(datar(i,j,lag))*SQRT(datadof(i,j,lag)))/$
        SQRT(1.-datar(i,j,lag)*2.))
         data_sig(i,j,lag) =
((datar(i,j,lag)*SQRT(datadof(i,j,lag)))/$
        SQRT(1.-datar(i,j,lag)*2.))  GT $
        T_CVF(.1,datadof(i,j,lag))
         datacomp(i,j,lag) = dataf(i,j,lag,0) +
dataf(i,j,lag,1)*tsval
      ENDFOR
     ENDFOR
   ENDFOR
```

Any suggestions will be greatly appreciated. This code was written nearly 2 years ago, so perhaps more recent versions of IDL handle this better?

Many thanks,
Charlotte
