Statistical tools in Climatology

René Garreaud

www.dgf.uchile.cl/rene
Gridded Analysis: set of maps with data on a regular grid (e.g., lat-lon, lon-lev)

- Spatial structure at a given time (displayed with isolines or colors)
- Time series at each grid point
Basic Operation with gridded data
Mean and Std Deviation

Average, standard deviation and other statistical parameters (max, min, skewness, etc.) are calculated using the time series at each grid point.
Example of mean fields...

Fig. 1. Austral summer (DJF) mean climatological fields of (a) SLP (contour interval is 2 hPa), (b) low-level (925–850 hPa) winds, and (c) upper-level (300–200 hPa) winds.
Use **compositing analysis** to describe the spatial structure of selected climate variables during the occurrence of events at your site.

Sometime, the events are associated to extreme values of a continuous time series. In this case, you can also use regression analysis (1Point correlation map), but the later method will fail if there is no a linear relationship...
Basic Operation with gridded data
Compositing Analysis (composite map)

Here one can use a univariate method (e.g., $t$-test) at each point to test for statistical significance of the anomalies.
Example of compositing analysis...time series

Fig. 3. (upper) The 30-min average of surface pressure, (middle) meridional wind, and (lower) solar radiation at station Lengua de Vaca (30.1°S, 71.5°W, 7 m MSL) from 27 Aug to 26 Sep 1994. In the upper panel, thin line is the observed surface pressure and thick line is the surface pressure with the mean diurnal cycle removed (anomaly filter). Crosses indicate selected coastal lows in this period (see text for further details).

Fig. 4. Composite traces of surface pressure (solid line, hPa) at six stations along the coast of Chile (station name and latitude is indicated at the top of the corresponding trace) based on 21 CL episodes during 1994. Composite trace based on 57 episodes during
Example of compositing analysis...spatial fields

Fig. 3. Convective index composite anomalies from day -2 to day +2. Contour interval is 10 W m$^{-2}$ for positive anomalies (solid contours) and 5 W m$^{-2}$ for negative anomalies (dotted contours). Black area indicates terrain elevations in excess of 2000 m. Light shading indicates the 95% significance level.
Cold minus Warm composites to enhance signal

Fig. 4. COLD – WARM composite SST during (a) DJF and (b) MAM. Contour interval is 0.2°C. Negative contours are dashed. Shaded areas indicate anomalies statistically different from zero (t-test) at 95% (heavy shade), 99% (medium shade), and 99.9% (light shade) significance levels.
Basic Operation with gridded data

1Point Correlation Map

Here one can use a univariate method at each point to test for statistical significance of correlation
r : Measure the covariability of two time series (-1 to +1)
r²: Fraction of variance in series Y explained by linear fit of series X
Annual mean Precip/SAT regressed upon index of large-scale modes (50 years of data)
How is the spatial pattern of SST when there is above/below normal Rainfall over the Altiplano?
Basic Operation with gridded data
Point-Point Correlation Map

Here one can use a univariate method at each point to test for statistical significance of correlation
Fig. 2. (a) Map of local correlation between monthly anomalies of precipitation and 300-hPa zonal wind, scale at the bottom. Dashed lines outline regions where annual mean precipitation exceeds 1000 mm yr$^{-1}$. (b) Same as in (a) but for local correlation between monthly anomalies of precipitation and 850-hPa zonal wind.
Large annual cycle
Small interannual variability

Small annual cycle
Large interannual variability
Intraseasonal noise?
Nice trend
Moderate annual cycle
Some noise…

Abrupt climate change?
Strong annual cycle?
Full time series
Mix regular cycles and interannual variability

Stgo. Rainfall

1-Point Correlation map
Stgo. Rainfall – 1000 hPa air temperature
Stgo. Winter Rainfall

1-Point Correlation map
Stgo. Winter Rainfall – 1000 hPa air temperature
Alternatively, use the anomaly filter, i.e. subtract annual mean cycle
Caution with seasonally dependent signal

Mean annual cycle

\[
x_{\text{time}} = x_{\text{year}}^{\text{month}} \quad N_{\text{time}} = 12 \times N_{\text{year}}
\]

\[
\bar{x}_m = \frac{1}{N_{\text{years}}} \sum_{y=1}^{N_y} x^y_m
\]

Anomalies

\[
x^y_m = x^y_m - \bar{x}_m
\]
Caution with seasonally dependent signal
1 Point Correlation Map between U300' over the Andes and Precipitation

DJF

JJA

All months
Most geophysical (met) variables exhibit some spatial coherence. Adjacent grid points covary with reference point...but how far?

U300 winter

Stgo Winter Precip
EOF / PC Analysis

- Powerful method to synthesize (large) data sets
- Allows to obtain the mode(s) that explain most of the variance in the data
- Each mode composed by a spatial pattern (EOF), time pattern (PC, loading factors) and fraction of variance accounted by
- It uses linear algebra: EOF/PC are linear combination of original data
- EOF = Empirical orthogonal functions (instead of analytical orthogonal functions)
EOF / PC Analysis

- User has to decide spatial domain (large-small)
- Pre-filter data (at least remove mean field)
- How many modes are retained (North’s rule)
- Each mode is orthogonal with the rest…not always true in nature
- Rotated EOFs / Extended EOFs
Leading EOF (dominant mode)

Let the data speak by itself

\(\lambda\): fraction of variance accounted by a mode

EOF Analysis
Fig. 8. (a) The first REOF of fractional cold ($T^*_c = 240$ K) cloud coverage $F^*$ in the central Andes. Partial EOF loadings $\geq 0.5$ for regions $\geq 1500$ m are shown and northeastern slope between 1500 and 3000 m contour line. (b)–(c) As in (a) but for REOFs 2–3.

Fig. 9. (a) The first RPC of DJFM fractional cold ($T^*_c = 240$ K) cloud coverage $F^*$. (b) As in (a) but for RPC 2. White symbols indicate average DJFM latitudinal position of 0 m s$^{-1}$ 200-hPa zonal wind component at 70°W (scale is reversed and plotted on right axis). (c) As in (a) but for RPC 3.
Fig. 13. (a) RPC 1 regressed upon DJFM 200-hPa NCEP–NCAR geopotential height and wind. Wind field is only plotted where the correlation between either zonal or meridional wind component and RPC 1 is significant at $p \leq 0.05$. Scale for wind vector (in m s$^{-1}$ per std dev) is shown below. Contour interval is 5 gpm per std dev; negative contours are dashed. (b) As in (a) but for NOAA-interpolated OLR and 200-hPa wind. Shading indicates OLR values above (below) 2.5 ($-2.5$) W m$^{-2}$ per std dev (see scale below). (c) As in (a) but for RPC 2. (d) As in (b) but for RPC 2.