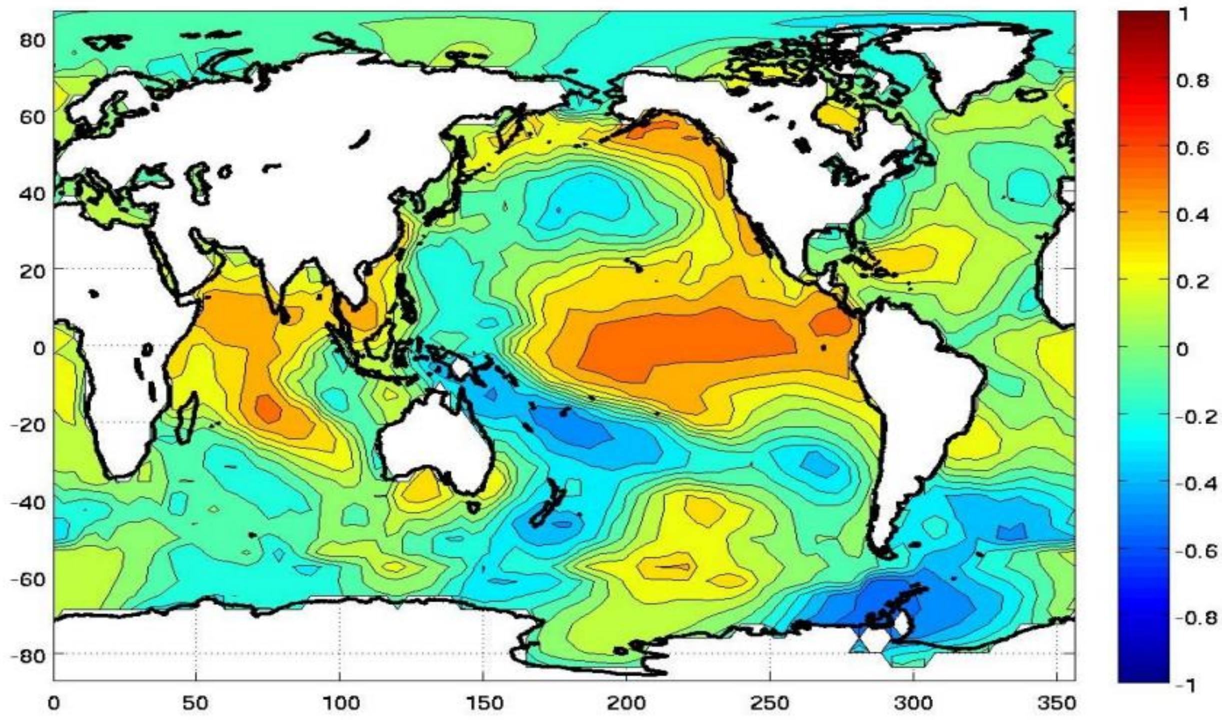
Análisis Estadístico de Datos Climáticos

Correlaciones y Regresión

2015

Correlaciones



- Mapa de correlacion entre anomalias de PP_artigas5102 y TSM en OND.

¿Como sé qué correlaciones son significativas?

Para la correlacion de Pearson existe el siguiente test.

- H0: r=0
- H1: r ≠ 0

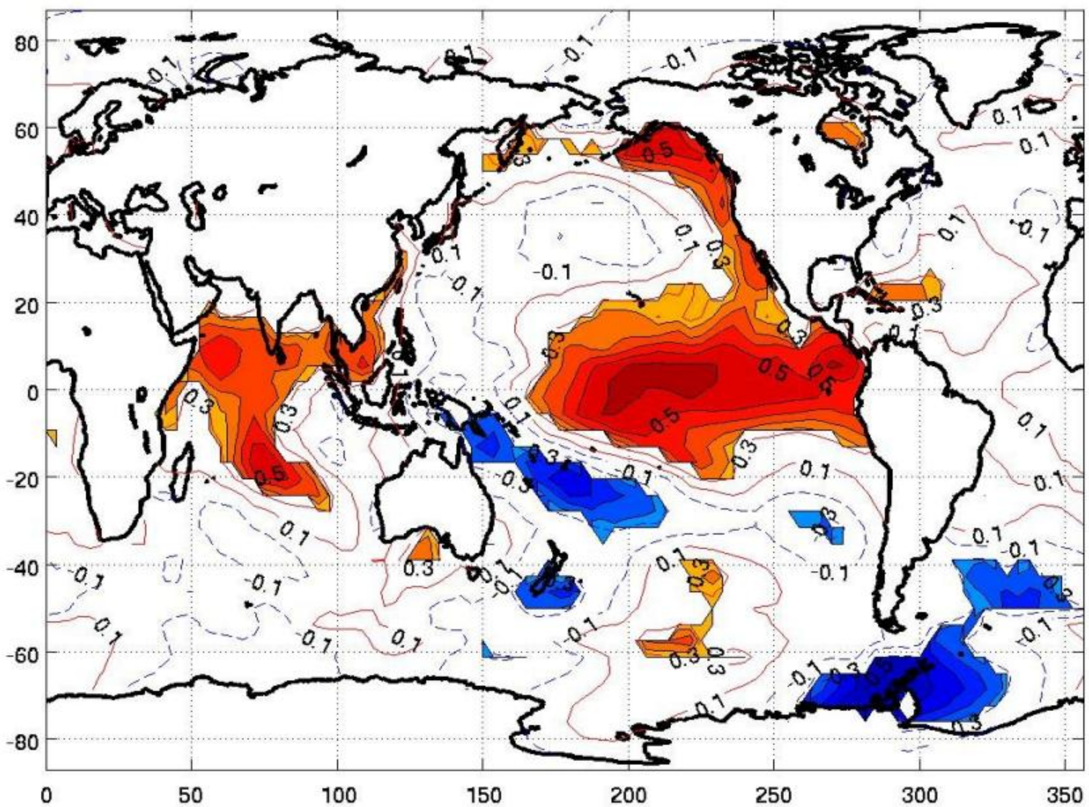
$$T = |(r)| \frac{\sqrt{(n-2)}}{\sqrt{(1-r^2)}}$$

- n es la longitud de la serie

- se compara T con valores criticos de la distribucion t de Student con n-2 grados de libertad

Este test requiere que ambas series sigan una distribución gausiana. Veremos más adelante una forma de resolver este problema.

Regiones de TSM con correlacion significativa con pp_artigas5102 en OND.



Se grafica en colores el estadístico T donde es significativo (0.05 y 2 extremos) y contornos de la correlación.

```
%Codigo Matlab
anomOND=(anom(10:12:end,:,:)+anom(11:12:end,:)...
+anom(12:12:end,:,:))/3; % anomalías de TSM

anompOND=(anomp(:,10)+anomp(:,11)+anomp(:,12))/3;% anom. precip en Artigas

%Calcula correlacion
for i=1:96
      for j=1:48
           correl(j,i)=corr(anompOND,anomOND(:,j,i));
      end
end

%Que correlacion es significativa?
%Test Ho: r=0 (compare to T distribution)
% 2 extremos al 5% de significancia
```

$$T = \left|(r)\right| \frac{\sqrt{(n-2)}}{\sqrt{(1-r^2)}}$$

```
tt=correl*sqrt(52-2)./sqrt((1-correl.^2));
jj=find(abs(tt)<=tinv(0.975,52-2)); tt(jj)=NaN;

contour_map(lon,lat,tt',0,(-5:0.5:5)) % grafica el estadístico T en colores (donde es significativo)
hold
contour_map(lon,lat,correl',0.2) % grafica contornos de correlación (en intervalos de 0.2)
```

Regresión

- Supongamos que quiero saber cual es la relacion entre el estado del Pacifico ecuatorial y la precipitacion en Sud América en OND.
Tomo Niño3.4 como indice.

- Una forma de ver esto es hacer composites, como ya hicimos.

- Otra forma es calcular un mapa de regresión: hacer una regresión lineal entre cada punto de grilla de precipitacion y Niño3.4. ➔

- **Observacion: Es muy útil trabajar con anomalías estandarizadas: $z=x/\sigma$.**

- **En la regresion**

$y = a + bx$

las unidades de [b]=(Unidades de y)/(Unidades de x)

En este ejemplo: [b]=mm/dia / °C

- **Si x está estandarizada, no tiene unidades y tiene desviación estandard=1**

- **Así, b se puede interpretar como la anomalía de y asociada (dependiendo del r) a una desviacion estandard de la variable independiente x.**

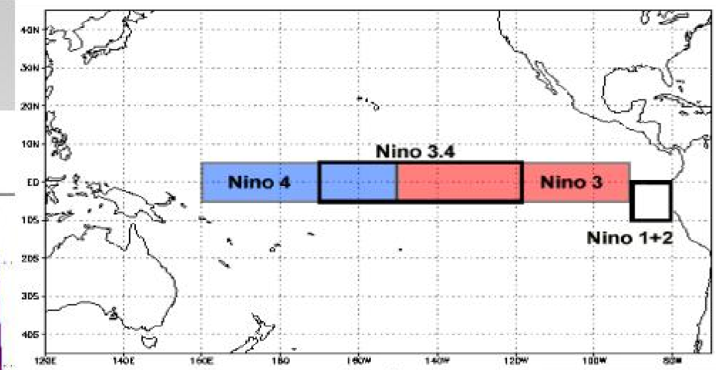
Nino3.4

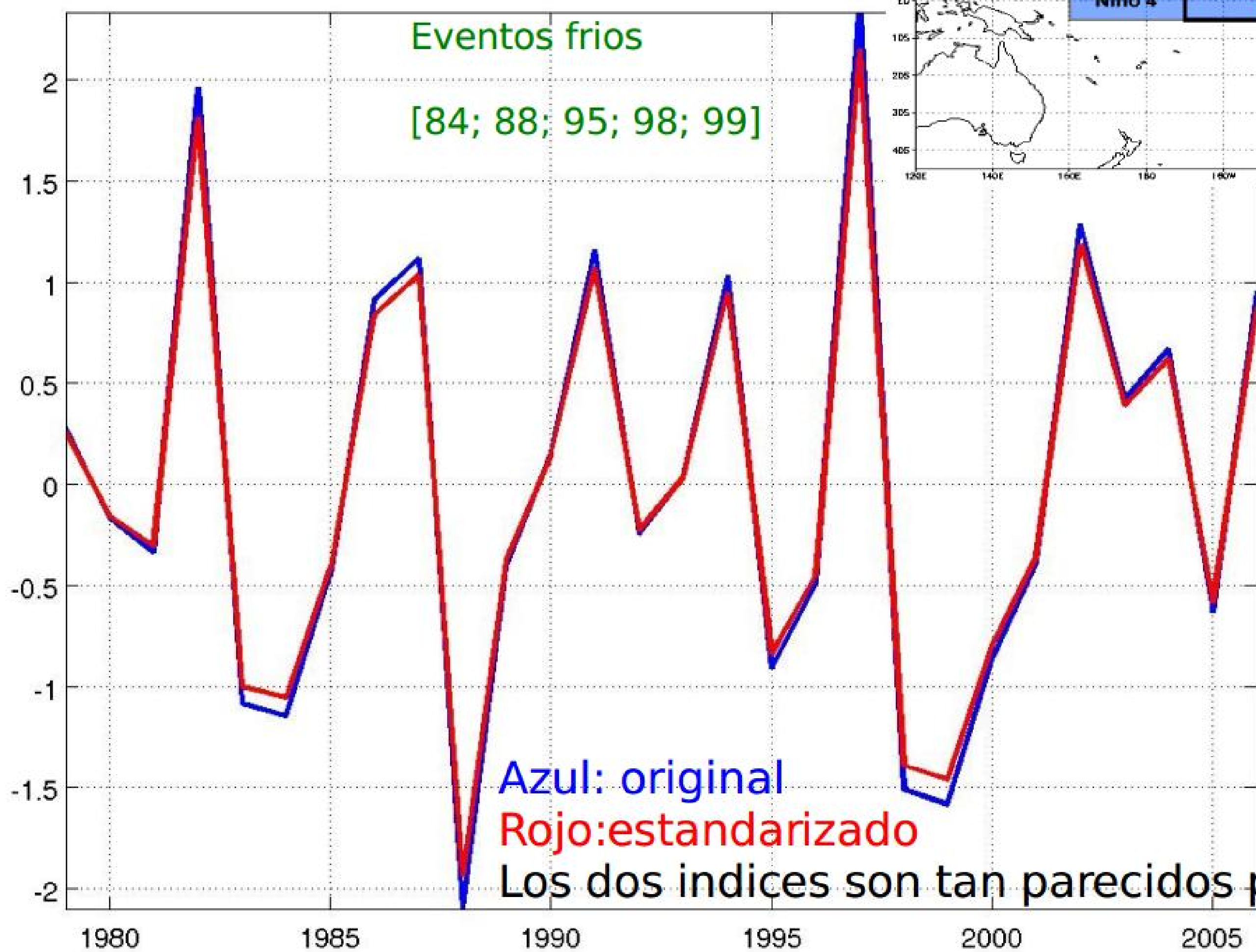
Eventos calidos

[82; 86; 87; 91; 94; 97]

Eventos frios

[84; 88; 95; 98; 99]

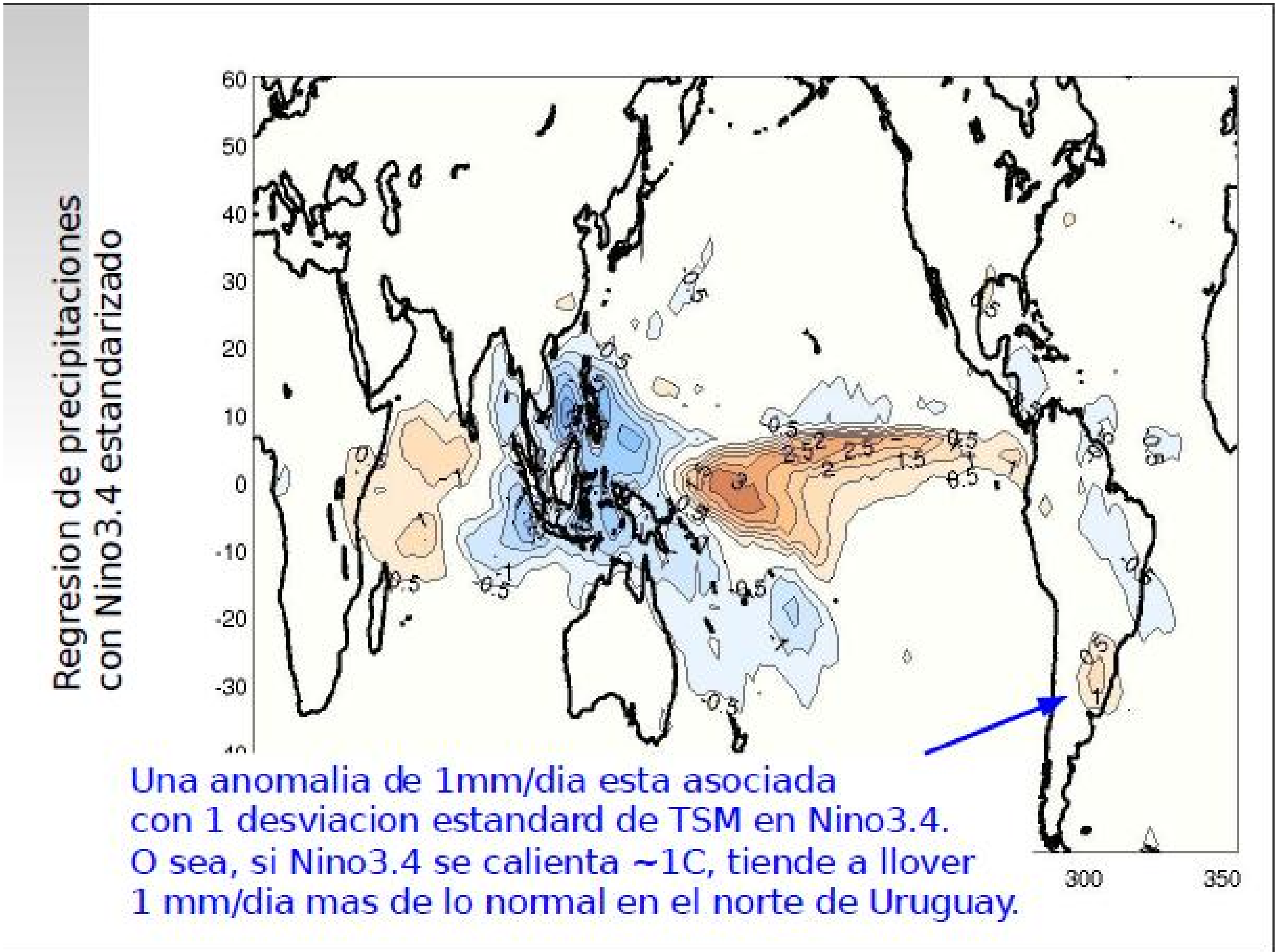




Azul: original

Rojo:estandarizado

Los dos indices son tan parecidos pues σ~1 C



Se grafica el coeficiente angular b.

Pero cómo sabemos en qué regiones es la regresión estadísticamente significativa?

- El coeficiente de multiple determinacion R^2

$$R^2 = \frac{SSR}{SST} = \frac{\frac{S_{XY}}{S_{XX}} S_{XY}}{S_{YY}} = \frac{S_{XY}{}^2}{S_{XX} S_{YY}} = \frac{(\sum (x_i - \bar{x})(y_i - \bar{y}))^2}{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2} = r^2$$

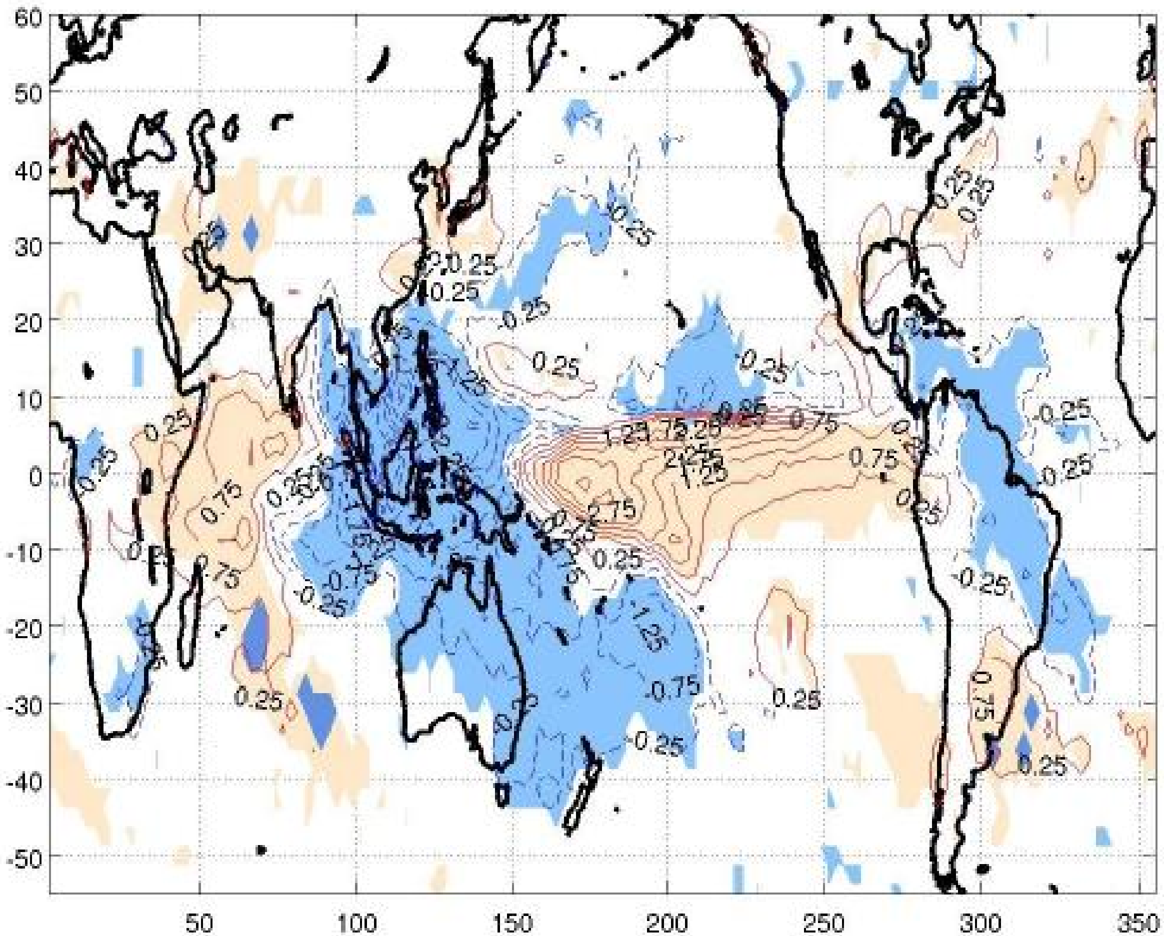
  y representa la habilidad de la recta estimada en representar las variaciones en los datos.

- Entonces para saber la significancia estadística calculamos el mapa de correlacion y aplicamos el test

$$T = |(r)| \frac{\sqrt{(n-2)}}{\sqrt{(1-r^2)}}$$

  comparando con la distribucion t.

Regresion y significancia estadistica al 5%



Se grafica en colores el estadístico T donde es significativo (0.05 y 2 extremos), y contornos del coeficiente angular b.

```
%Anomalias de Precipitacion
anompOND=(anomp(10:12:end,:,:)+anomp(11:12:end,:,:)+anomp(12:12
:end,:,:))/3;

%Anomalias de TSM
anomOND=(anom(10:12:end,:,:)+anom(11:12:end,:,:)+anom(12:12:end,
:,:))/3;

%Nino3.4 (es la gráfica que ya vimos)
figure
plot((1979:2006),nino34,'linewidth',2)
hold
nino34s=nino34/std(nino34);                    %Estandarizo el indice Nino34
plot((1979:2006),nino34s,'r','linewidth',2)
grid; axis tight

%Calculo Correlacion y Regresion
for i=1:144
    for j=1:72
        p=polyfit(nino34s',anompOND(:,j,i),1); b(j,i)=p(1);
        r(j,i)=corr(nino34s',anompOND(:,j,i));
    end
end
```

```
figure
%Que correlacion es significativa?
%Test Ho: r=0 (compare to T distribution)

tt=r*sqrt(28-2)./sqrt((1-r.^2));
jj=find(abs(tt)<=tinv(0.975,28-2)); tt(jj)=NaN;

% Contour estadístico (donde es significativo)
contour_map(X,Y,tt',0,(-20:20:20))
shading flat
hold

%Contour regresión
contour_map(X,Y,b',0.5)
axis([2 355 -55 60])
colormap(rednblue3)
caxis([-30 20])
```

Diferencias entre composites y regresión

• En un composite se toman los extremos y se comparan con los años “neutros”. El composite se puede hacer para extremos positivos y negativos y estos resultados no tienen por que ser opuestos (respuesta no lineal).

• En la regresion solo se considera la relacion lineal entre el predictando (y) y el predictor (x).