

Análisis Estadístico de Datos Climáticos

Composites

2015

Composites

El método de “composites” (o compuestos) consiste en clasificar los datos en categorías según una variable z , y comparar p. ej. los valores medios o anomalías de otra variable V según las distintas categorías de z .

Puede servir para identificar “señales” no muy fuertes que están ocultas debido a la existencia de “ruido”.

En otras palabras, la idea es formar conjuntos o categorías Θ de la variable o índice z , y estimar el valor esperado de la variable V , condicionado a que $z \in \Theta$.

Formalmente, el composite está dado por:

$$V_{\Theta} = \mathbf{E} (V_t \mid z_t \in \Theta) \quad (\text{t es un índice, que suele ser el tiempo})$$

En la práctica, esta expresión se sustituye por el estimador:

$$V_{\Theta} = \frac{1}{k} \sum_{j=1}^k V_{t_j} \quad \text{donde la suma se realiza sobre los tiempos } t_1, t_2, \dots, t_k \text{ para los cuales } Z_{t_j} \in \Theta.$$

Una forma de determinar si algunos aspectos de la señal capturada por z están expresadas en V es hacer una prueba de hipótesis a:

$$H_0: V_{\Theta_1} = V_{\Theta_2},$$

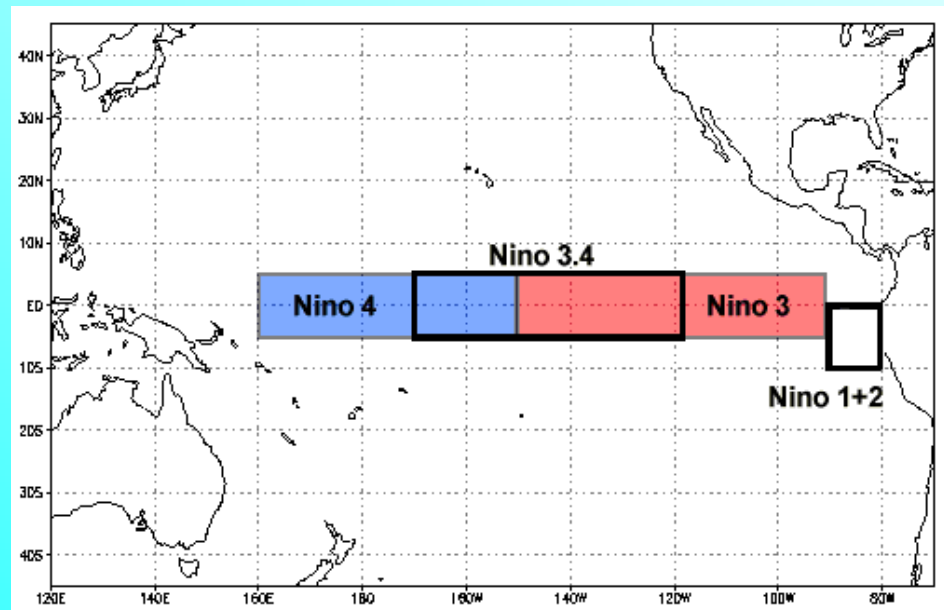
para conjuntos disjuntos Θ_1 y Θ_2 adecuadamente elegidos.

Esta prueba se suele hacer con un test de diferencia de medias que ya hemos visto.

(Von Storch y Zwiers, p. 378-380)

Composites

Ejemplo 1: componemos anomalías de **lluvias** globales en el trimestre OND según anomalías simultáneas de TSM positivas (eventos “cálidos”) o negativas (eventos “fríos”) en N3.4 en 1980-2000. (z es el índice de TSM en N3.4 y V son las precip en OND en América del Sur; los conjuntos Θ son los años cálidos, fríos, o neutros.)

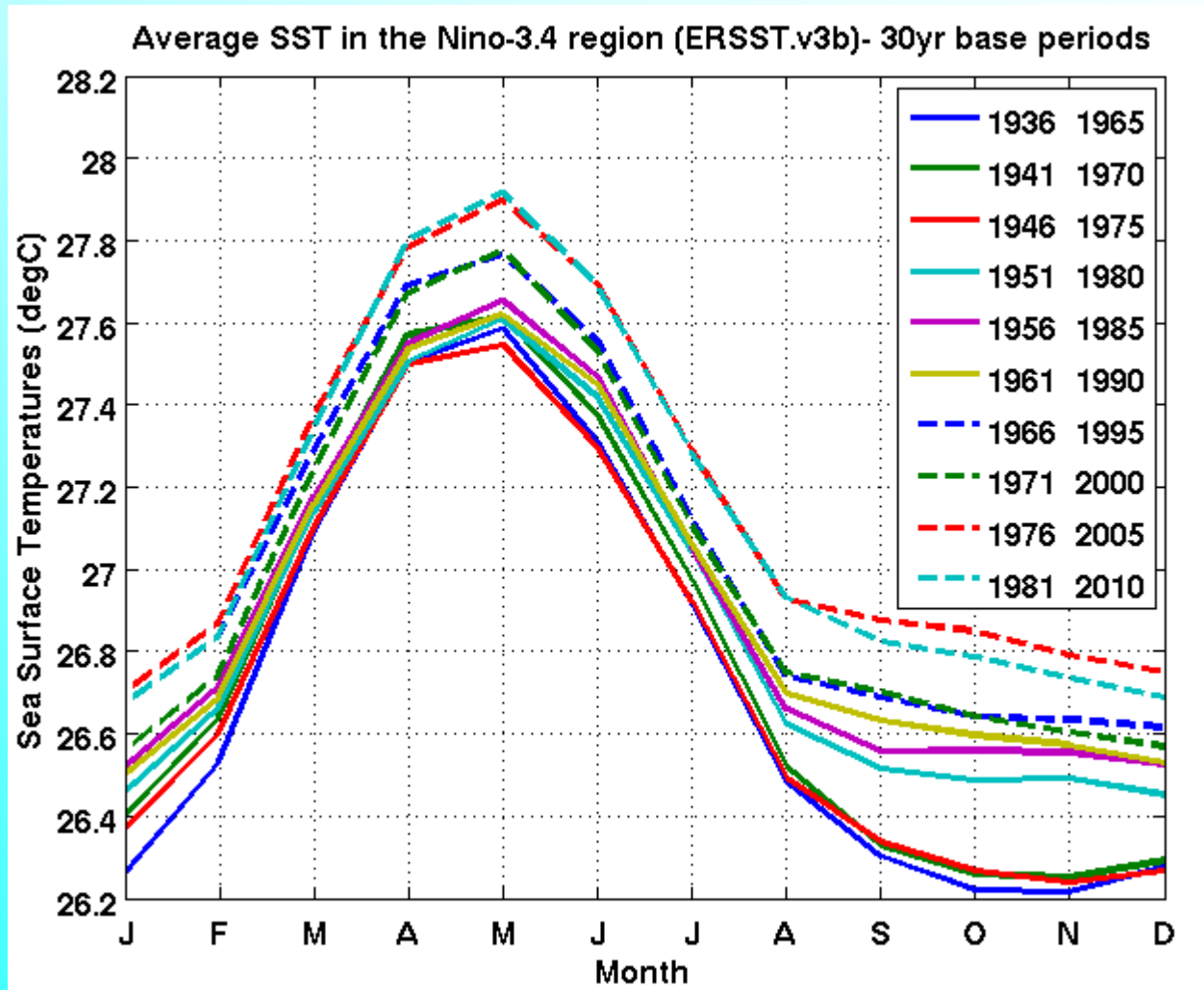


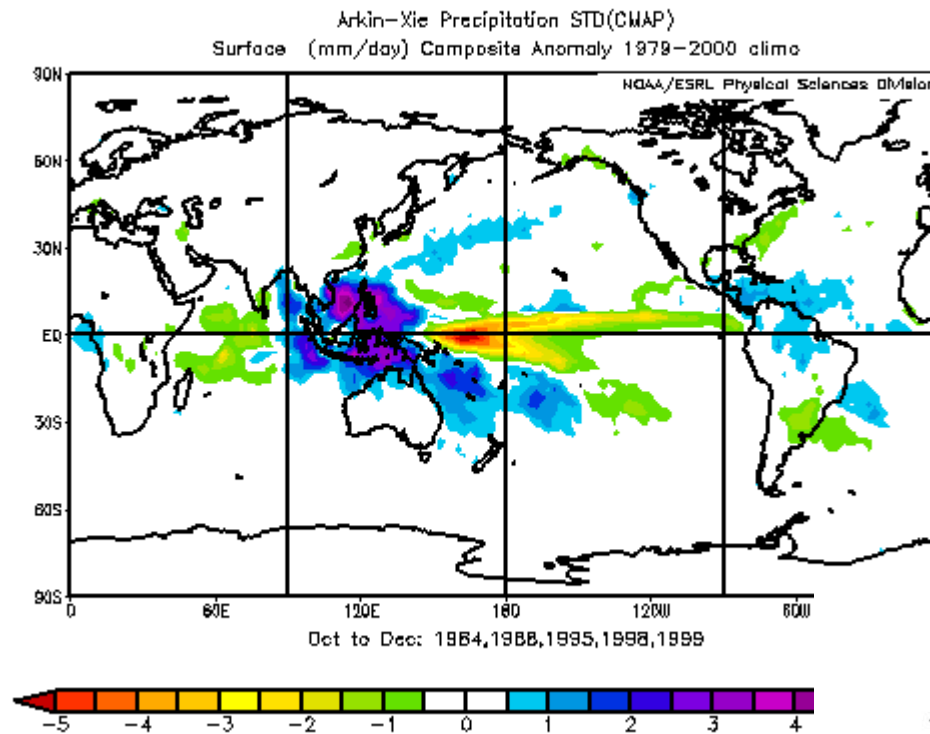
Episodios **cálidos** y **fríos** en la región N3.4 (1980-2000)

Year	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
1980	0.5	0.4	0.3	0.2	0.3	0.3	0.2	0.0	-0.1	0.0	0.0	0.0
1981	-0.2	-0.4	-0.4	-0.3	-0.2	-0.3	-0.3	-0.3	-0.2	-0.1	-0.1	0.0
1982	0.0	0.1	0.2	0.4	0.7	0.7	0.8	1.0	1.5	1.9	2.2	2.3
1983	2.3	2.1	1.6	1.3	1.0	0.7	0.3	-0.1	-0.5	-0.7	-0.9	-0.7
1984	-0.4	-0.2	-0.2	-0.3	-0.4	-0.4	-0.3	-0.2	-0.2	-0.6	-0.9	-1.1
1985	-1.0	-0.9	-0.8	-0.8	-0.8	-0.6	-0.6	-0.5	-0.6	-0.4	-0.4	-0.4
1986	-0.5	-0.5	-0.3	-0.2	-0.1	0.0	0.2	0.4	0.6	0.9	1.0	1.2
1987	1.2	1.3	1.2	1.1	1.0	1.2	1.5	1.7	1.6	1.5	1.2	1.1
1988	0.7	0.5	0.1	-0.3	-0.9	-1.3	-1.4	-1.2	-1.3	-1.6	-2.0	-2.0
1989	-1.8	-1.6	-1.2	-0.9	-0.7	-0.4	-0.4	-0.4	-0.4	-0.3	-0.2	-0.1
1990	0.1	0.1	0.3	0.3	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.4
1991	0.4	0.4	0.3	0.3	0.6	0.8	1.0	0.9	0.9	0.9	1.3	1.6
1992	1.8	1.7	1.5	1.4	1.2	0.9	0.5	0.2	-0.1	-0.1	0.1	0.3
1993	0.4	0.4	0.5	0.7	0.7	0.7	0.4	0.3	0.3	0.3	0.3	0.3
1994	0.2	0.2	0.2	0.3	0.4	0.4	0.5	0.5	0.7	0.9	1.3	1.3
1995	1.2	0.9	0.6	0.3	0.2	0.1	-0.1	-0.2	-0.5	-0.6	-0.8	-0.8
1996	-0.8	-0.7	-0.5	-0.3	-0.2	-0.2	-0.1	-0.2	-0.1	-0.2	-0.3	-0.4
1997	-0.4	-0.3	-0.1	0.3	0.8	1.3	1.7	2.0	2.2	2.4	2.5	2.5
1998	2.3	2.0	1.4	1.1	0.4	-0.1	-0.7	-1.0	-1.1	-1.2	-1.4	-1.5
1999	-1.5	-1.2	-0.9	-0.8	-0.8	-0.8	-0.9	-1.0	-1.0	-1.2	-1.4	-1.7
2000	-1.7	-1.4	-1.0	-0.8	-0.6	-0.6	-0.4	-0.4	-0.4	-0.5	-0.7	-0.7

http://www.cpc.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml

La NOAA ha cambiado la forma de cálculo de las anomalías en el Pacífico Ecuatorial

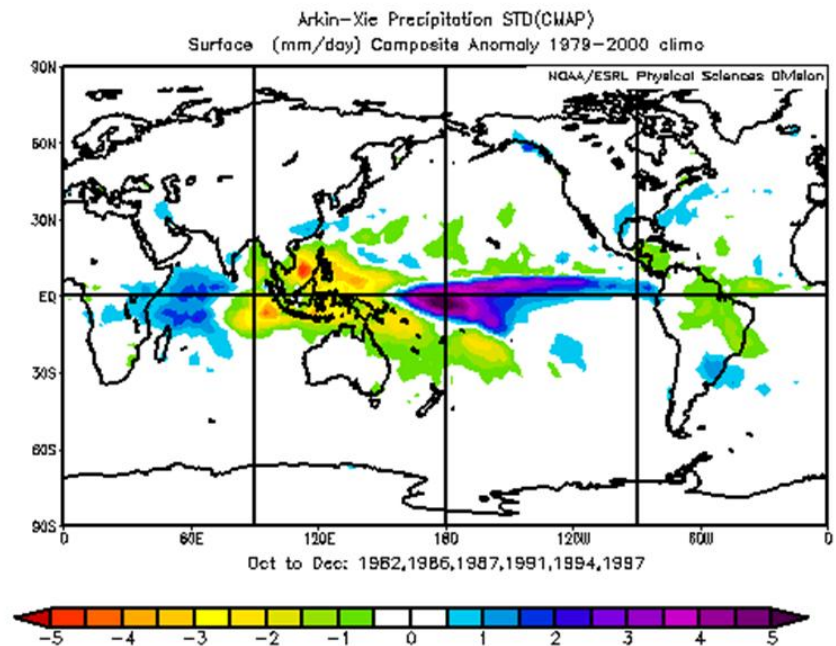




Compuestos de anomalías de TSM

Eventos fríos (1984-88-95-98-99)
(años Niña)

Eventos cálidos (1982-86-87-91-94-97)
(años Niño)



Observar que hay linealidad en las anomalías de TSM

Vamos a analizar si las anomalías de precipitación en OND en una región que cubre América del Sur, en los mismos años de eventos cálidos (N3.4) son estadísticamente significativas, con respecto a las de los años “neutros”.

Para ello se hace un test de Student (t-test) para comparar la diferencia entre las dos medias muestrales.

H_0 : tienen la misma media; H_1 : la media es diferente o...

Se asume que las muestras proceden de la misma población, tienen distribución normal (esto se puede verificar con el test de Lilliefors), son independientes y tienen igual desviación típica.

Comparo las medias de las muestras:

- Años Niño [82; 86; 87; 91; 94; 97] (estos son los t_k)**
- Años Neutros (los que no son Niños ni Niñas)**

- **El test es:**

$$\frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}} \rightarrow t_{n_1 + n_2 - 2}$$

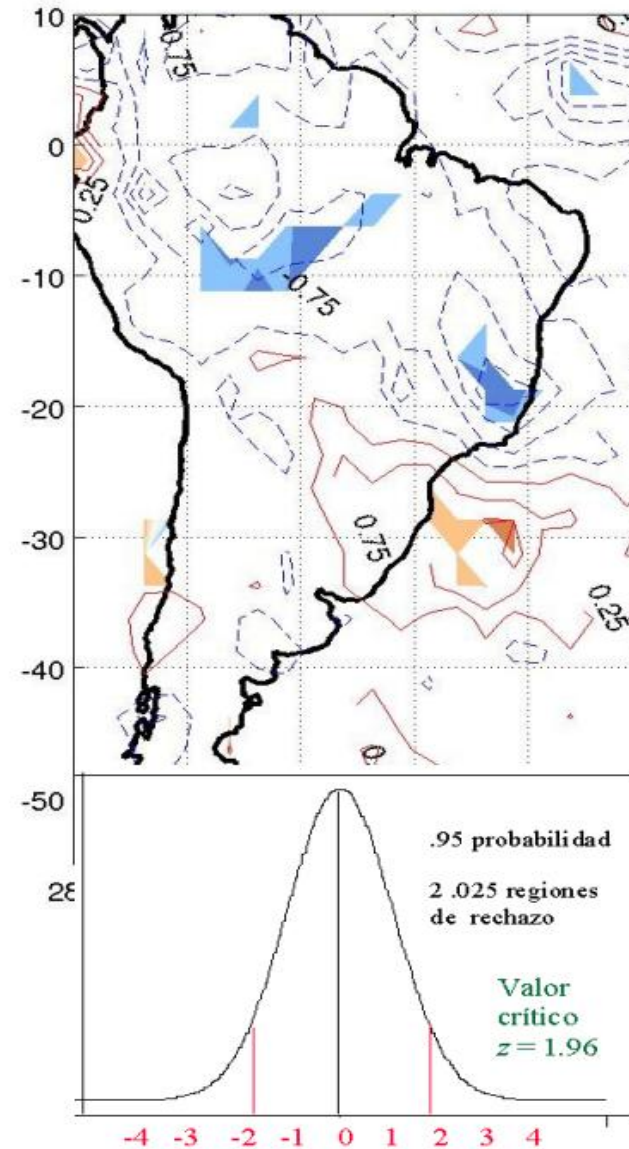
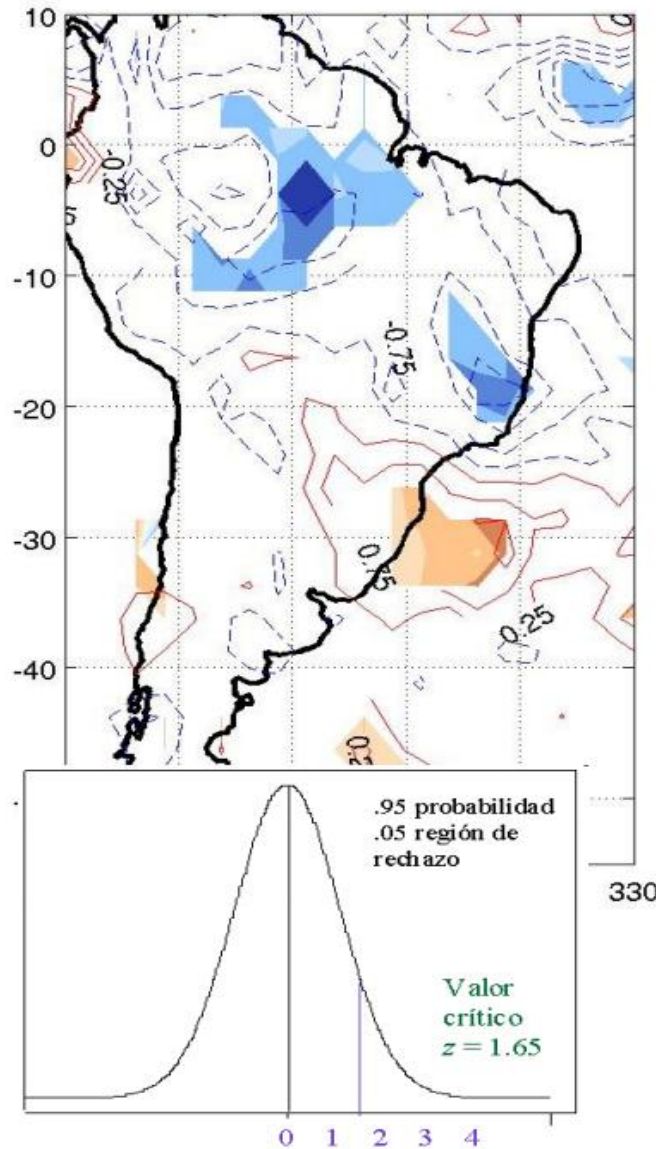
**(ya lo
vimos
con otra
notación)**

Trabajamos con datos grillados de precipitación (2.5° x 2.5°)

Anomalias significativas al 5%

1 extremo

2 extremos



Como tenemos menos información el test debe ser mas fuerte

(Si fuera la dist.normal)

```
%Matlab
```

```
[clim,anom]=climatology(pre,X,Y,0);
```

```
anomOND=(anom(10:12:end,::)+anom(11:12:end,::)+...  
          anom(12:12:end,::))/3;
```

```
%Años Niño y Niña
```

```
nino=[82; 86; 87; 91; 94; 97] 78;
```

```
nina=[84; 88; 95; 98; 99] - 78;
```

```
%Años Neutros
```

```
neutros=(79:106) - 78; neutros(nino)=NaN;  
neutros(nina)=NaN;
```

```
neutros=neutros(~isnan(neutros))';
```

%%% t-test

$$\frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}} \rightarrow t_{n_1 + n_2 - 2}$$

Los cálculos se hacen en cada punto de la grilla de precipitación

```
anomONDnino=mean(anomOND(nino,:::));
```

```
anomONDneut=mean(anomOND(neutros,:::));
```

```
nino_neu=anomONDnino - anomONDneut; % numerador
```

```
dof=length(nino)+length(neutros) - 2; % grados de libertad
```

```
sp=sqrt((var(anomOND(nino,:::))*(length(nino) - 1)+...  
var(anomOND(neutros,:::))*(length(neutros) - 1) )/dof);
```

```
tt=nino_neu./(sp*sqrt(1/length(nino)+1/length(neutros)));
```

```
tt=squeeze(tt); tt2=tt;% tenemos el valor del estadístico en cada punto
```

%-- nivel 5% (test de 1 extremo)

`jj10=find(abs(tt2)<=tinv(0.95,dof)); % identifica los puntos no significativos`

`tt2(jj10)=NaN.*ones(size(jj10)); % pone NaN en esos puntos`

`subplot(1,2,1)`

`contour_map(X,Y,tt2',0,(5:1:5)); % grafica el tt2 donde es significativo`

`hold; shading flat;`

`contour_map(X,Y,squeeze(nino_neu)',0.5); % grafica la diferencia de medias con paso de 0.5 entre los contornos`

`colormap(rednblue3); caxis([-6 5]); axis([280 330 -55 10])`

%-- nivel 5% (test de 2 extremos)

```
jj5=find(abs(tt)<=tinv(0.975,dof));
```

```
tt(jj5)=NaN.*ones(size(jj5));
```

```
subplot(1,2,2)
```

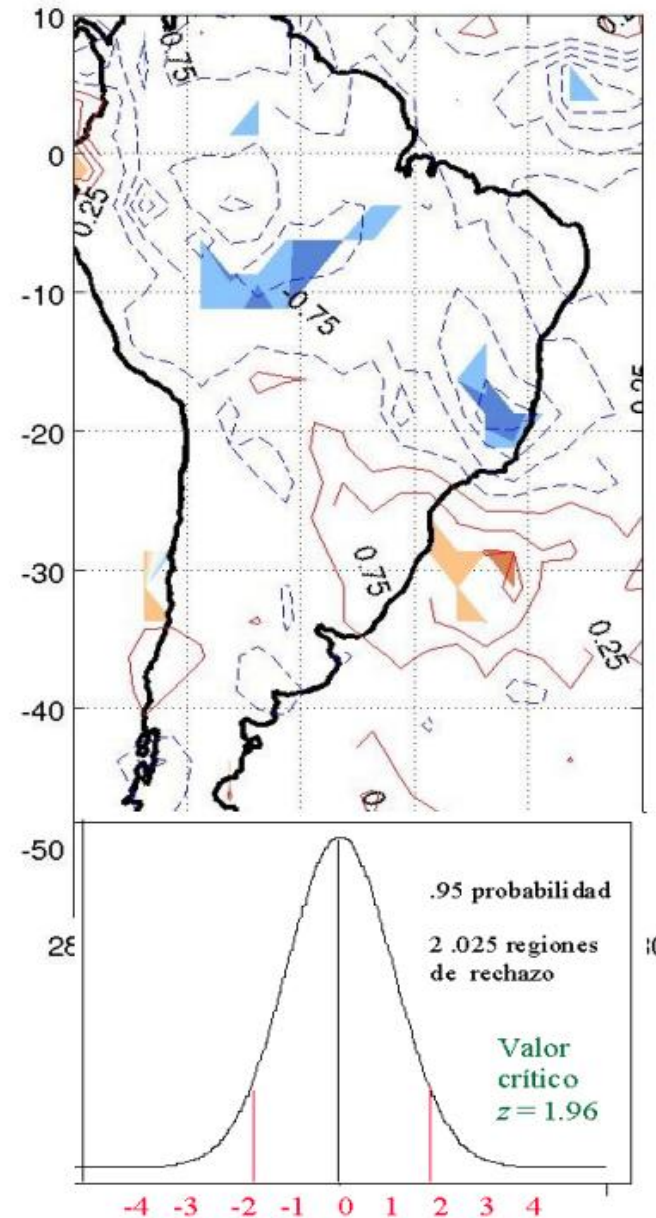
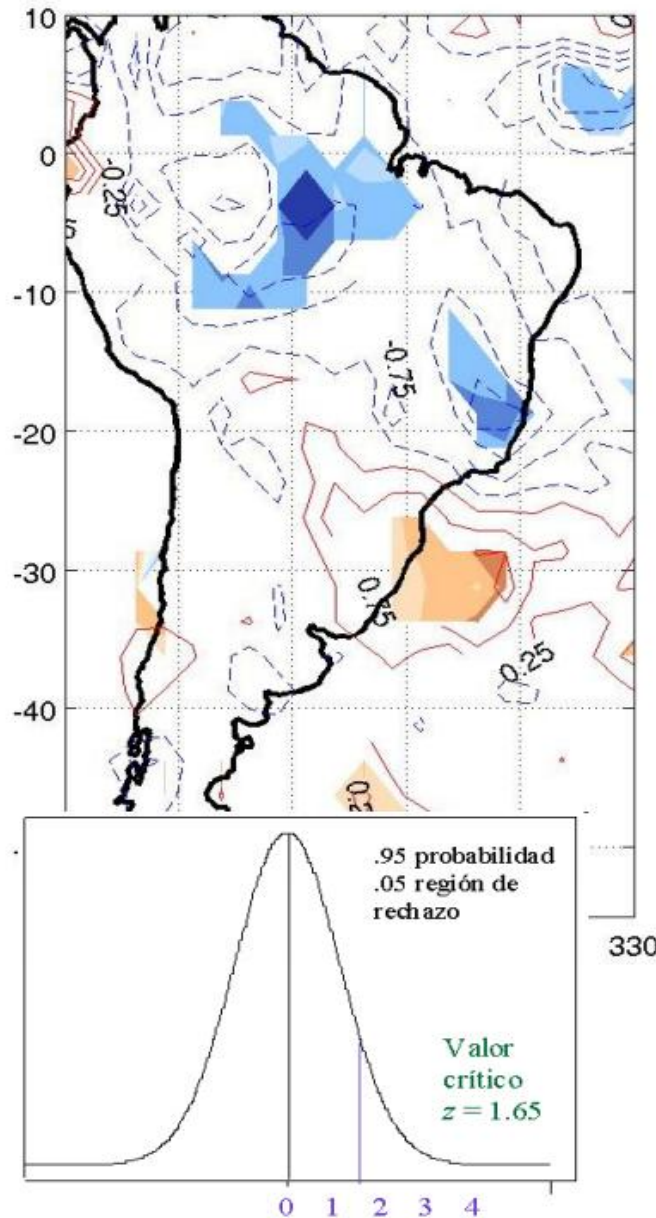
```
contour_map(X,Y,tt',0,(-5:1:5)); hold;  
shading flat
```

```
contour_map(X,Y,squeeze(nino_neu)',0.5);  
colormap(rednblue3); caxis([-6 5]); axis([280 330 -55 10])
```

Anomalías significativas al 5%

1 extremo

2 extremos



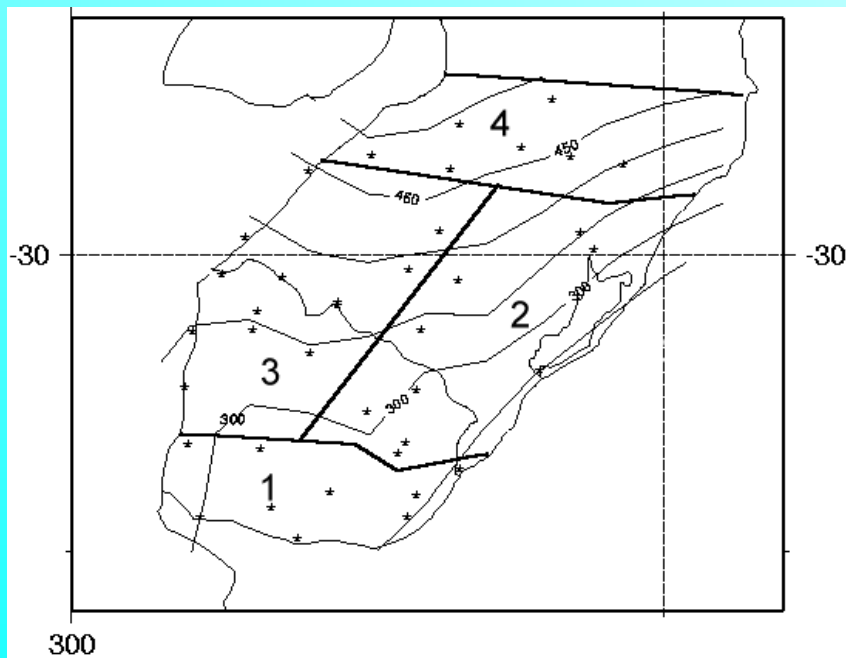
Como tenemos menos información el test debe ser mas fuerte

(Si fuera la dist.normal)

Composites

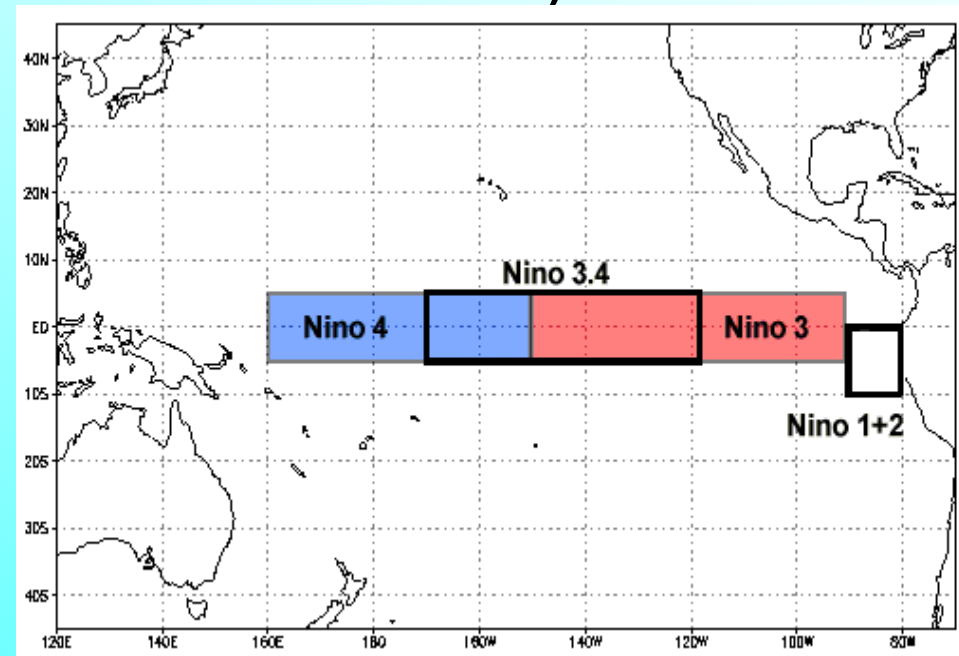
Ejemplo 2: aplicación al pronóstico

4 regiones en Uruguay-Río Grande
do Sul
(21 + 17 pluviómetros en 1950-1998)
OND



Región Niño 3.4 en el
Océano Pacífico
Jul-Ago anterior

(el pronóstico se quería emitir en set.
2004)

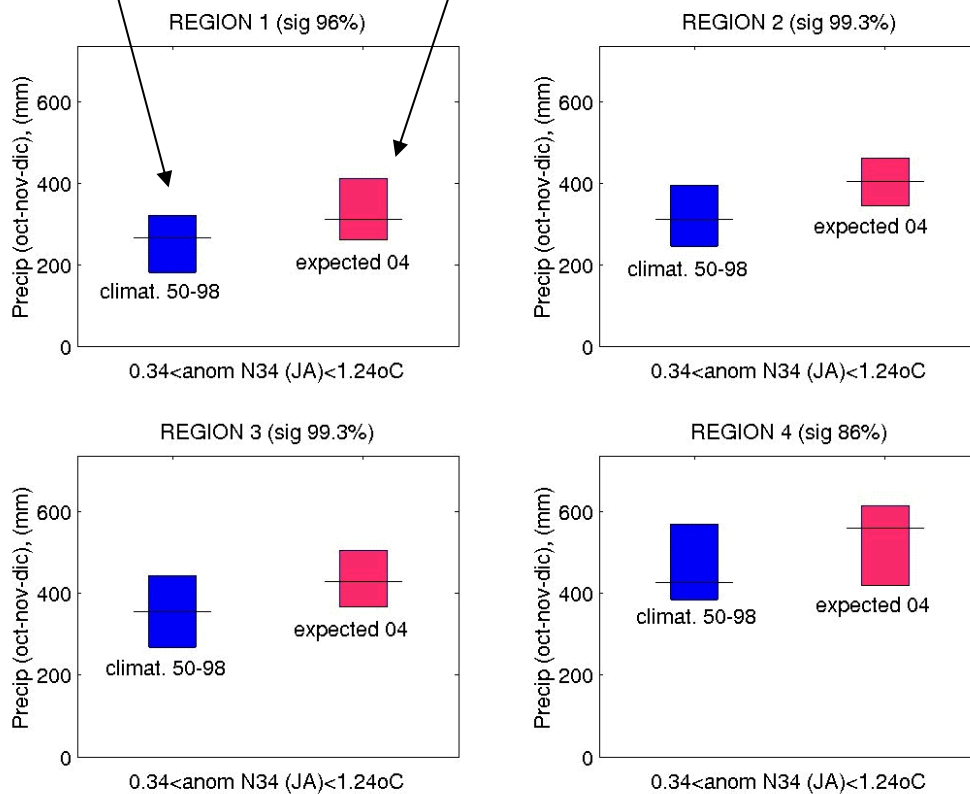


Composites

Climatología de precipitación en OND

Subpoblación condicionada a $(0.34 \text{ }^{\circ}\text{C} < (\text{Jul-Ago N3.4}) < 1.24 \text{ }^{\circ}\text{C})$

(situación similar a la de N3.4 en Jul-Ago 2004)



Estos resultados también deben someterse a pruebas para determinar si son estadísticamente significativos. Ej: Mann-Whitney