Composite Analysis Instructions

Last revised 6.10.2005

This document covers how to 1) create station composite analysis spreadsheets for various data types (i.e., MeanT, Precip, etc.), 2) compute historical station composites analyses based on El Niño/La Niña, 3) apply risk analysis techniques, and 4) produce a current composite analysis forecast based on El Niño/La Niña.

This exercise uses Microsoft Excel as the platform for analysis. Excel is used for training purposes to provide a more detailed explanation of the process and methodology employed. The statistical software SPLUS will be utilized operationally and for developmental composite analysis at the local office. Although SPLUS streamlines the process through a graphical user interface for data input and statistical analyses, the methodology will not be evident to the user. It is important to develop skill in the composite analysis methodology through Excel first.

Composite Analysis Instruction Section Outline

1   Climatological Data
2   Compute Historical Composite Analysis
3   Trend Adjustment
4   Risk Analysis
5   Composite Analysis Forecast
6   Creating Graphical Products
7   Hindcast Forecast
8   Composite Analysis Forecast Verification
9   Resources

Assumptions

This document assumes the use of Microsoft Excel 2000 or beyond (refer to http://www.nws.noaa.gov/om/csd/pds/pcu2/res/students/primerExcel.htm).
A basic understanding of statistics (probabilities, distribution tails, statistical significance) is assumed (refer to http://www.nws.noaa.gov/om/csd/pds/pcu2/res/students/primerStat.html).

Data Files and Templates

The following data files are provided at _____:

    Composite_XXX_MMMYYYY_cccc_template.xls
    oni.xls

Alternatively, the ONI data (numerical data only, no episode text) is available at:
    http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml

The CPC Niño 3.4 consolidated SST forecast is available at:
    http://www.cpc.ncep.noaa.gov/products/precip/CWlink/ENSO/sstcon34.txt

Definitions

Category – refers to either above, near, or below normal for a specific climate variable

CPC consolidated Niño 3.4 SST forecast – SST in the Niño 3.4 region for 13-3 month leads (JFM, FMA, MAM, etc.).

Composite analysis – is a sampling technique based on the conditional probability of a certain event occurring (e.g., El Niño or La Niña)

Conditional probability – is the probability that a given event will occur if it is certain another even has taken place or will take place. It is based on the Probability Distribution Function (PDF) representing only situations satisfying a certain condition (like the existence of a strong El Niño).

Niño 3.4 region – located in the Pacific Ocean between 5°N and 5°S and 120° and 170°W; see:
http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/nino_regions.html

Oceanic Niño Index (ONI) – the SST departure from the average SST in the Nino 3.4 region. ONI is based on a 3-month running mean and homogeneous historical SST analyses (1905-1998). ONI utilizes the NOAA operational definition of El Niño and La Niña:

    El Niño $T_{dep} \geq +0.5\ ^{o}C$
    La Niña $T_{dep} \leq -0.5\ ^{o}C$
    Neutral $-0.5\ ^{o}C < T_{dep} < +0.5\ ^{o}C$

Terciles – the division of data into 3 equal parts and referred to as above, near, or below.
Statistical Glossary - http://meted.ucar.edu/climate/climanom/glossary.htm#n

Section 1: Climatological Data Last revised 6.8.2005

Retrieving Climatological Data

1. Download the data set from xmACIS. Go to: http://xmacis.rcc-acis.org or for a specific Weather Forecast Office (WFO) add '/XXX' where XXX is the 3 letter station identifier, for example: http://xmacis.rcc-acis.org/OKX

2. Select area of interest. Fig. 1.1 shows the WFO list to choose from.
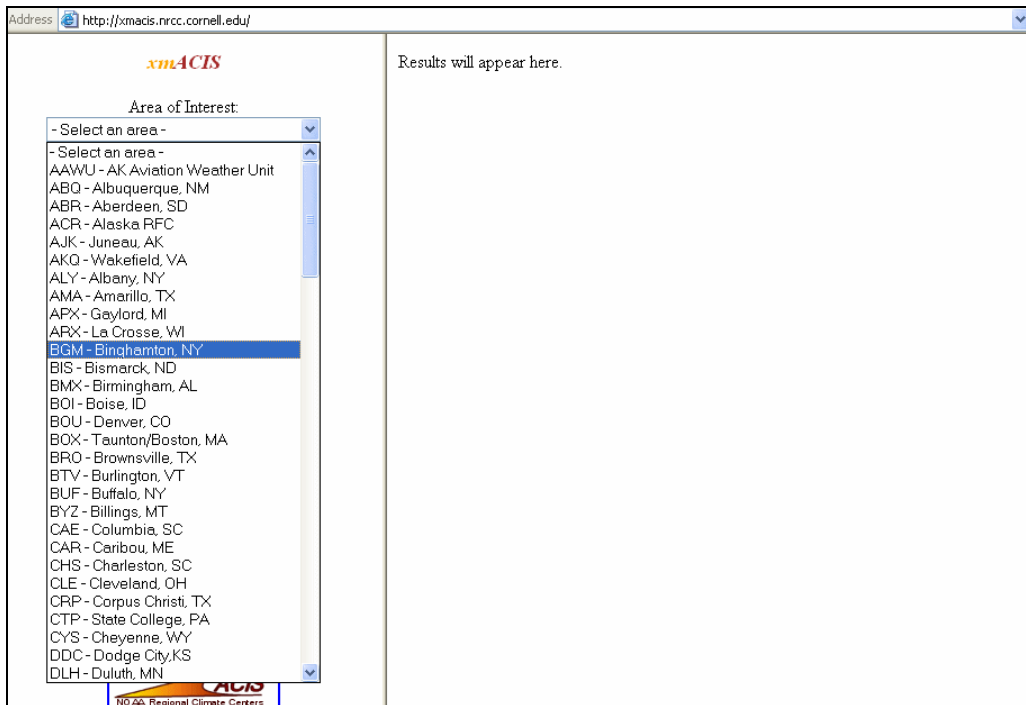
Fig. 1.1 xmACIS showing the *"Area of Interest"* list of WFOs.

3. Select Routine. Select *"Monthly Avgs/Totals"* as the routine (Fig. 1.2).

4. Select the climate variable and period of interest. Fig. 1.3 shows the different climate variables to select. You should select the variable *"Avg Temperature".*

The start and end year are defined by the user (Fig. 1.4).
Enter:
  Start Year (1950)
  End Year (current year)

5. Submit the request.

6. Data display. Data requested for a station will display on right side of display (Fig.1.5). **<u>*Note*</u>**: check the years of the data that are displayed. Some station's data may

not be available starting in 1950. So when copying the data into a spreadsheet, in later steps, make sure you copy the data into the correct location (correct starting year).

Fig. 1.2. xmACIS showing the list of routine choices.
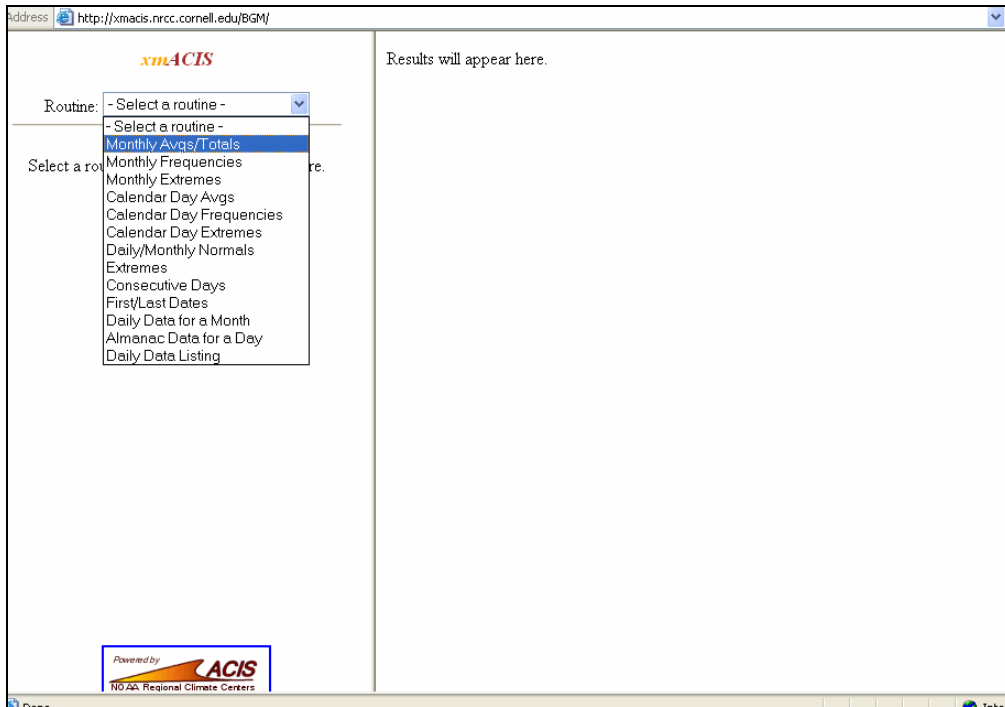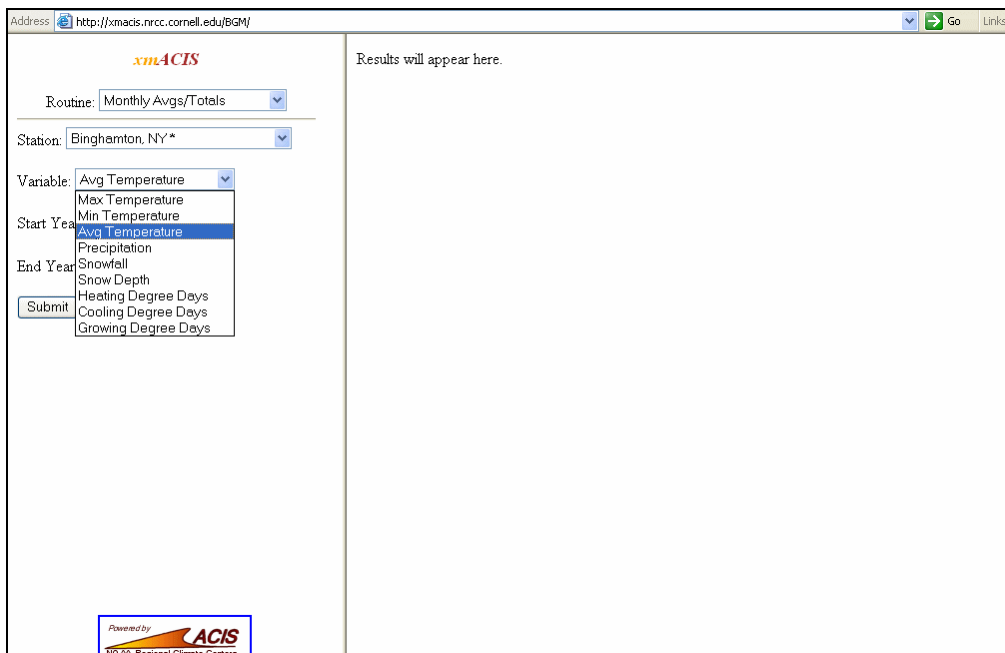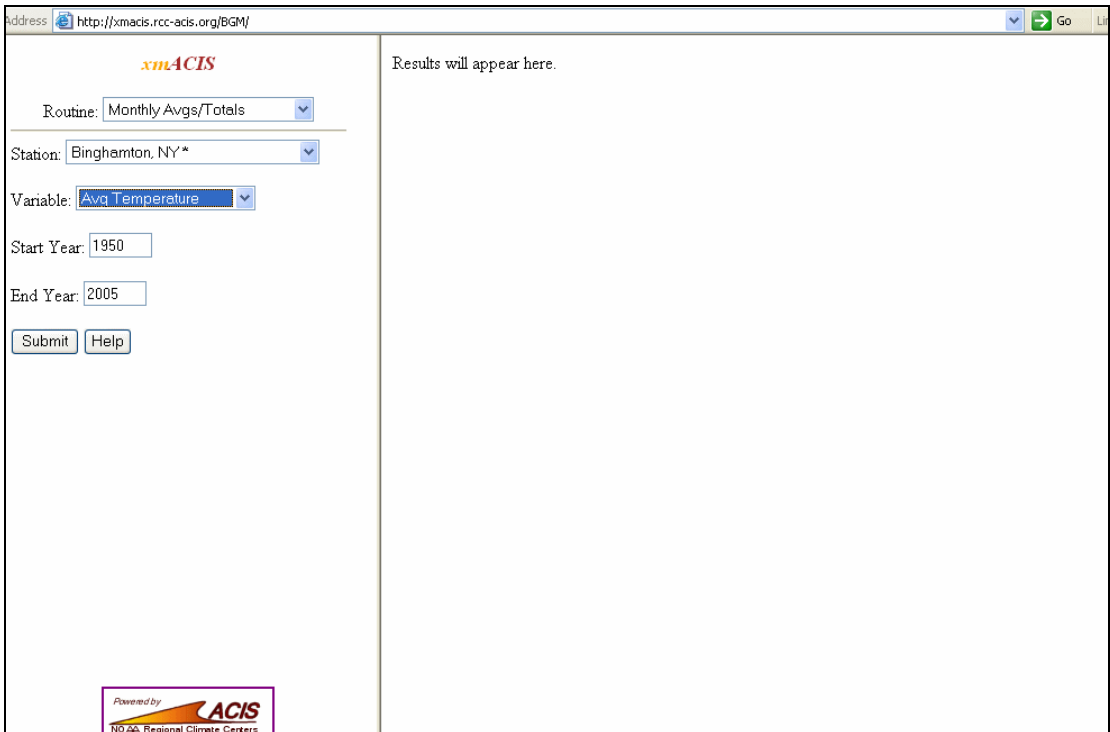

Fig. 1.3 xmACIS showing the variable selected and selection choices.

Fig. 1.4 xmACIS showing the variable selected and *"Start and End Year"* input areas.


```
BINGHAMTON EDWIN A LINK FIELD (KBGM)
Monthly Totals/Averages
Average Temperature (degrees F)
Years: 1951-2005
```

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
1951	–	–	–	–	–	–	69.1	66.0	59.1	51.1	32.1	27.1	50.7
1952	25.1	25.6	30.3	47.3	51.8	65.5	72.1	66.8	60.9	44.9	39.5	28.7	46.5
1953	28.2	27.5	33.6	41.8	57.1	65.5	69.4	67.6	61.0	51.5	41.0	31.8	48.0
1954	19.8	30.7	29.6	47.1	53.2	65.3	67.6	65.1	59.5	52.3	37.0	25.1	46.0
1955	19.6	23.9	31.6	48.4	57.9	64.3	74.5	71.7	59.1	51.3	34.5	19.7	46.4
1956	21.0	25.0	26.7	40.6	50.9	65.1	65.5	66.1	55.8	50.3	38.0	31.0	44.7
1957	17.3	26.9	32.6	45.4	54.0	67.8	67.5	64.6	60.5	47.0	39.1	30.6	46.1
1958	21.2	16.9	30.2	44.6	51.7	58.6	68.3	67.6	59.6	48.2	39.5	19.6	43.8
1959	21.5	23.1	30.8	46.6	58.9	65.3	71.3	71.6	64.3	51.1	36.1	30.5	47.6
1960	25.7	26.1	22.7	48.2	56.7	64.5	66.8	68.0	62.2	48.6	40.6	20.8	45.9
1961	16.7	25.6	32.0	39.8	52.3	64.5	68.2	66.8	65.9	53.2	37.3	25.0	45.6
1962	19.8	18.8	31.0	44.3	59.1	64.5	66.0	66.9	56.9	49.2	34.7	22.8	44.5
1963	19.2	16.6	35.4	45.5	53.5	65.3	67.0	62.8	54.7	54.8	41.0	19.3	44.6
1964	24.0	18.8	31.9	43.0	58.4	61.9	70.6	65.2	61.2	48.3	42.9	27.9	46.2
1965	19.8	22.7	28.7	40.3	59.1	63.1	66.6	66.6	60.8	46.8	36.9	30.4	45.1
1966	19.0	22.8	33.1	41.5	50.5	65.8	71.1	69.4	56.8	46.9	41.1	26.5	45.4
1967	29.0	18.4	29.4	43.6	47.6	68.2	67.8	66.2	59.1	48.6	32.9	28.9	45.0
1968	17.1	17.3	34.4	48.4	51.3	62.9	69.4	67.0	61.8	49.9	36.7	23.5	45.0
1969	22.1	23.0	29.9	46.9	56.3	63.8	67.0	67.5	60.1	47.9	37.8	23.5	45.5
1970	16.5	25.0	28.9	45.0	57.0	62.8	67.9	67.2	60.6	50.9	39.2	24.2	45.4
1971	15.6	23.6	28.4	40.7	54.3	66.2	65.9	64.0	62.8	55.4	35.7	30.8	45.3
1972	24.6	20.6	28.1	40.4	58.8	62.2	71.5	65.8	59.8	42.8	33.5	29.4	44.8
1973	25.8	20.1	39.9	45.4	51.5	66.3	70.2	69.6	59.7	52.6	39.8	29.0	47.5
1974	27.2	22.2	32.0	48.7	52.5	63.2	69.7	68.7	56.9	44.9	37.3	27.8	45.9
1975	24.9	24.4	27.2	37.0	62.0	66.5	72.8	67.4	56.5	51.4	44.6	26.3	46.7
1976	16.8	30.3	35.5	47.0	52.8	67.0	65.7	65.5	57.3	43.7	31.0	20.2	44.4
1977	12.0	24.1	38.0	47.3	59.0	61.8	69.9	66.5	60.3	46.6	40.2	24.5	45.9
1978	19.1	15.0	27.6	41.5	58.5	63.7	68.4	69.5	58.3	47.6	39.2	27.9	44.7
1979	21.1	13.6	37.9	42.5	54.9	62.1	69.0	65.6	58.7	48.4	42.6	31.2	45.6
1980	23.5	20.2	31.8	46.5	58.8	62.0	70.4	70.6	62.2	44.9	33.7	21.3	45.5
1981	15.2	30.7	32.3	47.5	57.2	65.2	69.4	67.8	58.4	46.3	38.2	26.5	46.2
1982	14.9	24.4	32.3	42.2	59.7	62.1	69.9	64.3	61.0	50.6	42.1	34.3	46.5

Fig. 1.5. xmACIS showing the results for BGM Monthly Totals/Averages for Avg Temperature from 1950-2004. Note: BGM's data begins in 1951 not 1950.

Moving the Data to the Station Spreadsheet

1. Set up an Excel spreadsheet.
    a. Open a new Excel spreadsheet.

    b. Save this spreadsheet as Data_XXX.xls, where XXX is a three letter identifier for the station you are investigating (Use *"File…Save As"* in the menu).

    c. Right mouse click on the Sheet 1 tab at the bottom and select *"Rename"* from the pop up menu. Name the tab for the data that will be copied into this sheet (e.g., MNTN [Mean Monthly Temperature] the four letter climate variable abbreviation assigned by NCDC).

2. Copying the data into the spreadsheet.

    Copy the data, 1950-present, (*Control-C*) including column headings from xmACIS. Paste (*Control-V*) it all into the page you named in the previous step in cell A1. **_Note_:** check the years of the data that are displayed. Some station's data may not be available starting in 1950. So when copying the data into a spreadsheet, in later steps, make sure you copy the data into the correct location. For example, if the data starts with 1951 data, copy the data into the correct starting year, 1951 not 1950. Additionally, instructions below assume starting with 1950 data. If your data starts with a different year, adjust instructions accordingly.

3. Formatting the data.
    a. From the Data menu, click on *"Text to Columns…",* this arranges the spreadsheet data into columns. Use the default values and click *"Next "* twice and then *"Finish"*. This will redo the spreadsheet with each data column in a separate column in your spreadsheet.

    b. Save your work and <u>save often</u>.

    c. Delete* the Annual column. When you are done, you should have the Year, Jan, Feb, Mar, etc. in consecutive columns in the spreadsheet.

    *To delete unneeded columns, *left -click* on the letter above the column, which will highlight that column. Hold down the *Control* key to highlight additional columns to be deleted. When you have several highlighted, *right-click* in one of the highlighted blue areas and select *"Delete"* from the popup menu. If you accidentally delete a column you didn't intend, from the Edit menu select *"Undo"* to restore the lost data. Continue to delete columns until you have deleted all the unnecessary columns after the Year column.

Creating the Composite Analysis Spreadsheet:

4. Setup the composite analysis spreadsheet.
<mark>Template is available at:</mark>

a. Open the template, Composite_XXX_MMMYYYY_cccc_template.xls. Save this spreadsheet as Composite_XXX_MMMYYYY_cccc.xls, where XXX is a three letter identifier for the station you are investigating, MMM is the 3 letters of the month the forecast is issued, YYYY is the year, and cccc is the four letter climate variable abbreviation assigned by NCDC (alternatively, a four letter identifier you create).

b. Open up the three month period composite analysis sheet (e.g., JFM) that you wish to examine. This spreadsheet contains 13 labeled sheets (JFM, FMA, MAM, AMJ, MJJ, JJA, JAS, ASO, SON, OND, NDJ, DJF, JFM).The three letter identifier represents a 3-month period (e.g., JFM for January-February-March). Rename the tab with the correct cccc identifier, in this example, MNTN, so the tab name should now be JJA MNTN.

You should now have two spreadsheets open – the one with the xmACIS data correctly formatted in it (station data spreadsheet) and the appropriate station and variable composite analysis spreadsheet with the three month sheet (composite analysis spreadsheet) for the period you want to investigate.

5. Copying station data to composite analysis spreadsheet.
See Fig. 1.6 for what the Composite_XXX_MMMYYYY_cccc.xls will look like.
a. In Data_XXX.xls, click on the data set tab you will copy (e.g., MNTN). Copy the Year column starting with 1950 (left mouse click to highlight the Year column, then, *right click* to bring up the menu. Select *"copy"*. Your selection will have dashed lines around it and be shaded blue.

b. Paste the Year column into Composite_XXX_MMMYYYY_cccc.xls in cell A2.

b. Place the cursor in Column E in the last cell with a value. Grab the lower right corner and pull down to the present year. This transfers the formulas used for the 3-month mean.

c. Place the cursor in cell B56. Grab the lower right corner and pull down to the present year. This transfers the formatting. Repeat for cell C56 and D56.

d. In Data_XXX.xls. Copy the portion of the data table from the station spreadsheet for the three month period you are looking at from 1950-present. (Highlight the data from 1950-present for the three months you wish to examine using the *left mouse button*, then *right click* to bring up the menu.

Select *"copy"*. Your selection will have dashed lines around it and be shaded blue.

e. Paste the data into Composite_XXX_MMMYYYY_cccc.xls in the appropriate 3-month sheet in cell B2 (to the right of the year column). We will now be working solely in this newly named compositing spreadsheet.

f. All values are in red. **LEAVE THEM THIS WAY!** The cells are formatted and will change as you complete the rest of the instructions.

6. Calculate a 3-month weighted average of the climate variable.
Note: In the template, this is already formatted! For Leap Year, see the note below.
a. For each year, the 3-month weighted average should be calculated as follows:

= [(Month 1 x number of days) + (Month 2 x number of days) + (Month 3 x number of days)]   (1)
                        total number of days in the 3 month period

b. In Excel the formula is shown below for the 3-month weighted average for 1950 in cell E2 (from =). *Only do this if you are creating the spreadsheet from scratch.*

Excel 3 month weighted average = ((B2*nd)+(C2*nd)+(D2*nd))/total nd   (2)

where "nd" is the number of days for that month.

Note: For February, use nd=28, but during a leap year use nd=29. This needs to be changed manually for leap years.

7. Add Oceanic Nino Index (ONI) data.
See Fig.1. 6 for what the Composite_XXX_MMMYYYY_cccc.xls will look like.
The ONI data is the 3-month temperature with an event episode, El Niño, Neutral, La Niña, as defined by NOAA's definition of El Niño/La Niña.

a. Open ONI.xls
i. On the ONI tab, copy the data for the 3-month period from 1950 to present into cell F2.
ii. On the ONI_word tab, copy the event descriptors for the 3-month period from 1950 to present into cell G2.

b. Columns F and G should be formatted based on the NOAA operational definition of El Niño and La Niña. You may need to format only additional

years that were added what is included on the template. To format any additional cells, follow these steps to format

    i. Select column F (ONI data).
    ii. From the Format menu, select Conditional Formatting.
    iii. For condition 1: cell value is ≤ -0.5 and format Blue
       For condition 2: cell value is ≥ +0.5 and format Red
    iv. Select column G (Episode).
    v. From the Format menu, select Conditional Formatting.
    vi. For condition 1: cell value is La Niña and format Blue
       For condition 2: cell value is El Niño and format Red

Table 1. NOAA operational definition of El Niño and La Niña based on sea surface temperature (SST) departures in the Niño 3.4 region.

NOAA operational definition of El Niño and La Niña
El Niño $SST_{dep} \geq +0.5°C$
La Niña $SST_{dep} \leq -0.5°C$
Neutral $-0.5°C < SST_{dep} < +0.5°C$

	A	B	C	D	E	F	G	H
1		APR	MAY	JUNE	AMJ	ONI	episode	
2	1950	48.1	58.4	70.4	58.96044	-1.4	La Nina	
3	1951	53.6	62.9	69.9	62.14176	0	neutral	
4	1952	54.8	60.8	74.1	63.20659	0.2	neutral	
5	1953	52.3	63.2	73.6	63.03516	0.6	El Nino	
6	1954	53.9	60.3	71.7	61.94835	-0.6	La Nina	
7	1955	53.3	65.2	70	62.85934	-0.6	La Nina	
8	1956	48.4	58.9	72.4	59.88901	-0.6	La Nina	
9	1957	53.2	63.4	74.5	63.6967	0.5	El Nino	
10	1958	52.7	59.1	67.5	59.75934	0.5	El Nino	
11	1959	53.4	65.1	71.6	63.38571	0.1	neutral	
12	1960	54.1	62.6	72.2	62.96264	0.1	neutral	
13	1961	48.6	58.6	71.5	59.55604	0.2	neutral	
14	1962	51.5	63.6	72.7	62.61099	-0.3	neutral	
15	1963	52.2	60.6	70.9	61.22637	-0.1	neutral	
16	1964	49.7	65.6	71.3	62.23736	-0.8	La Nina	
17	1965	50.7	67.1	72.5	63.47363	0.4	neutral	
18	1966	48.6	59	74	60.51648	0.3	neutral	
19	1967	49.7	55	71.8	58.79121	-0.3	neutral	
20	1968	54	59.4	69.2	60.85055	-0.3	neutral	
21	1969	53.9	62.9	70.9	62.57033	0.6	El Nino	
22	1970	51.1	62.6	70.5	61.41319	0	neutral	
23	1971	49.6	59.4	72.1	60.35604	-0.8	La Nina	
24	1972	47.6	59.7	65.8	57.72198	0.6	El Nino	
25	1973	53	58.9	72.7	61.5044	0.5	La Nina	

Fig. 1.6. Composite_XXX_MMMYYYY_cccc.xls showing the 3-month, ONI data, and episode.

Section 2: Compute Historical Composite Analysis

Creating the Tercile Information

See Fig. 2.6 for what the Composite_XXX_MMMYYYY_cccc.xls will look like.

1.   **Copy data from 1971-2000.**

   a. First, copy the years (1971-2000). To do this, highlight cell A23 and drag the cursor to 2000 year cell (cell A52). With the years 1971-2000 highlighted, right click on the highlighted area and select copy from the pop-up menu. Move the cursor to cell I5, right click in cell I5 and select paste from the pop-up menu.

   b. Copy the data from 1971-2000. Left click on cell B23 and highlight all the data (each month and the 3-month) to present year cell. *Right click* and choose *"copy"*.

   c. Remove colored formatting. In column J, highlight cells I5 to I34. From the *"Format"* menu, select *"Conditional Formatting"*. Click *"Delete"*. Check Condition 1 and Condition 2. Click *"OK"*. Click *"OK"*, again. Repeat this for column K, cells K5 to K34; column L, cells L5 to L34; and column M, cells M5 to M34.

2.   **Determine the tercile break points.**

   a. Highlight the break points for each month and 3-month average. The tercile information we are looking for are the average of the 10th and 11th slots (lower tercile) and of the 20th and 21st slots (upper tercile). The slot position values are located in column N.

   Check if for any missing data. If no data are missing proceed. Otherwise, if there are missing data, see Table 1 in the Missing Data box below to determine the slots to average to establish the terciles.

      i. Highlight the cells from I5 though cell M34.

      ii. Click on the "*Data*" menu and select "*Sort…*".

      iii. In the dialogue box that pops up, make sure "*No header row*" and "*Ascending*" are selected. Then choose what column to Sort by, choose column J and click "*OK*".

      iv. Select cells J14 and J15 (the 10th and 11th slots). *Right click* on these highlighted cells, and select "*format cells*". Click on *"Font"* tab, under *"Font*

Style" select Bold and under "*Color"* select the dark blue color. Click "*OK*". Those two cells should change color to dark blue.

v. Select cells J24 and J25 (the 20th and 21st slots). *Right click* on these highlighted cells, and select "*format cells"*. Click on "*Font"* tab, under "*Font Style"* select Bold and under "*Color"* select the red color. Click "*OK*". Those two cells should change color to red.

vi. Reselect the entire tercile table from I5 through M34. Select *Data…Sort* and sort using column K. Make sure the "*No header row*" and "*Ascending"* is selected. Select cells K14 and K15 and color these values blue. Select K24 and K25 and color these values red.

vii. Reselect the entire tercile table from I5 through M34. Select *"Data…Sort"* and sort using column L. Make sure the "*No header row*" and *"Ascending"* is selected. Select cells L14 and L15 and color these values blue. Select L24 and L25 and color these values red.

viii. Reselect the entire tercile table from I5 through M34. Select *"Data…Sort"* and sort using column M. Make sure the "*No header row*" and *"Ascending"* is selected. Select cells M14 and M15 and color these values blue. Select M24 and M25 and color these values red.

ix. Reselect the entire tercile table from I5 through M34. Select *"Data…Sort"* and sort using column I to place this table back in its original order. Make sure the "*No header row*" is selected.

b. Calculate the divisions between lower, middle and upper terciles.
i. Select cell J2 and place the average between the two blue values located in this column. This is the lower tercile boundary. Use the true average of the two values including any 'halves' – **do not round**.

ii. Select cell J3 and place the average between the two red values in that column.

iii. Do the same for K2 and K3, L2 and L3, and M2 and M3 using the same technique (results shown in Fig. 2.6).

Missing Data

When there are *no missing data*, the data are sorted for each month and 3-month season in ascending order. Then the 10th and 11th slots for the lower tercile value and 20th and 21st slot for the upper tercile value are averaged respectively. With missing data, the slots that are averaged change.

Some data, especially for the most recent year, may show up as -99999. If you

have access to these data, via a regional climate center or xmACIS, you will want to replace the missing data with actual values if available.

What to do for missing data?

After copying your data into the tercile table (1971-2000), sort each month column, as normal, into ascending order. The missing data will appear at the bottom of this sort (as   -9999, ####, M, or however you indicate the missing data). These data will not be used.

Table 1 shows the slots to average to determine the lower and upper tercile values with missing data. In some cases, you use one slot value as the tercile value, instead of averaging two. For example, with only 29 years of data, you will use slot 10 as the lower tercile value and slot 20 for the upper. **If you have less than 21 years of data, <u>DO NOT</u> perform composite analysis for that season, as this is not enough data to make the composite analysis reliable.**

Table 1. Determining the upper and lower terciles with missing data.

Years of Available Data	Slot Positions to Average for Lower Tercile			Slot Positions to Average for Upper Tercile	
21	7	8		14	15
22	7	8		15	16
23	8	8		16	16
24	8	9		16	17
25	8	9		17	18
26	9	9		18	18
27	9	10		18	19
28	9	10		19	20
29	10	10		20	20
30	10	11		20	21

Courtesy of Paul Hamilton, CWSU Freemont (formerly NWS Western Region HQ, Hydrology and Climate Services Division)

c. Tercile break points highlighted.
Note: In the template, this is already formatted!
   i. The data table from 1950-present (cell B2-E55 or so) will now show values in the lower tercile colored blue, the middle tercile colored black, and the upper tercile colored red. This is done using conditional formatting which are built into this spreadsheet.

   To manually set the conditional formatting, select the data to be formatted (cell B2-present year cell), then from the *"Format"* menu, select

"Conditional formatting". Set the conditions and formatting as follows in Table 2:

Table 2. Conditional formatting information to manually format data 1950-present.

	Condition 1	Format		Condition 2	Format
Month1 (cell B2-present year cell)	≤ cell J2	Blue, bold		≥ cell J3	Red, bold
Month2 (cell C2-present year cell)	≤ cell K2	Blue, bold		≥ cell K3	Red, bold
Month3 (cell D2-present year cell)	≤ cell L2	Blue, bold		≥ cell L3	Red, bold
MMM (cell E2-present year cell)	≤ cell M2	Blue, bold		≥ cell M3	Red, bold

	H	I	J	K	L	M	N
1			APR	MAY	JUNE	AMJ	
2		lower	51.55	61.25	70.65	61.53242	
3		upper	53.05	63.35	72.6	62.57308	
4							
5		1971	49.6	59.4	72.1	60.35604	1
6		1972	47.6	59.7	65.8	57.72198	2
7		1973	53	58.9	72.7	61.5044	3
8		1974	54.2	59.9	68.9	60.98791	4
9		1975	47.1	64.3	69.7	60.40989	5
10		1976	55.6	61.5	74.1	63.70879	6
11		1977	52.6	64.1	69.4	62.05604	7
12		1978	50.3	59.6	70	59.96264	8
13		1979	49.9	63	68.9	60.62637	9
14		1980	53.2	64.7	70.1	62.68901	10
15		1981	54.3	63.8	72.7	63.6022	11
16		1982	50.2	63.3	66.8	60.13516	12
17		1983	51.7	58.9	72.5	61.00989	13
18		1984	50.3	61.3	73.6	61.72857	14
19		1985	53.9	64.6	68.8	62.45714	15
20		1986	53.1	65.4	71.5	63.35604	16
21		1987	52.7	63.3	73	63.0033	17
22		1988	50.8	62.2	72.1	61.70549	18
23		1989	51.4	62.5	72.9	62.26923	19
24		1990	53	59.7	72.4	61.67802	20
25		1991	55.6	69.1	74.8	66.52857	21

Fig. 2.6. Composite_XXX_MMMYYYY_cccc.xls showing the tercile information.

Interrogating the Data

3.   Sort the 1950-present table by episode.
a. Highlight from cell A2 through G54 (or most current year). On the *"Data"* menu click "*Sort"*. Click the "*No header row*" option and sort by column G. Click "*OK*" and the data will sort by episode. (Fig. 2.7)

b. Highlight the El Niño events.
i. Count the number of El Niño events in the below (blue), normal (black), and above normal (red) categories for each separate month and for the three month season. For example, in April for El Niño there are 3 below, 6 neutral, and 5 above.

Composite Analysis Instructions Section 2   5

ii. Write these numbers in the appropriate cells J38 to M40 (Fig. 2.8, blue boxed area). The total event column will automatically fill in as you enter data into the table.

iii. Make sure that the total number of events for each El Niño/Neutral/La Niña episode is the same for each month and 3-month season.

Tip: An easy way to count the number of occurrences is to use the left mouse button and highlight the first cell to be counted (e.g., all the below normal events). Continue holding down the mouse button as you drag the cursor to the last cell of those events. Above column A at the top of the spreadsheet, you will see a white box with something like 5R x 1C. The 5 in this case means that you have highlighted 5 below normal events. This technique will be VERY useful when you are computing neutral cases which have over 30 events possible.

c. Highlight the La Niña events.
   i. Count the number of La Niña events in the below (blue), normal (black), and above normal (red) categories for each separate month and for the three month season. For example, in April for La Niña there are 9 below, 2, near, and 6 above.

   ii. Write these numbers in the appropriate cells J43 to M45 (Fig. 2.8). The total event column will automatically fill in as you enter data into the table.

   iii. Make sure that the total number of events for each El Niño/Neutral/La Niña episode is the same for each month and 3-month season.

	A	B	C	D	E	F	G
1		APR	MAY	JUNE	AMJ	ONI	episode
2	1950	52.3	63.2	73.6	63.03516	0.6	El Nino
3	1951	53.2	63.4	74.5	63.6967	0.5	El Nino
4	1952	52.7	59.1	67.5	59.75934	0.5	El Nino
5	1953	53.9	62.9	70.9	62.57033	0.6	El Nino
6	1954	47.6	59.7	65.8	57.72198	0.6	El Nino
7	1955	50.2	63.3	66.8	60.13516	0.7	El Nino
8	1956	51.7	58.9	72.5	61.00989	1	El Nino
9	1957	52.7	63.3	73	63.0033	1	El Nino
10	1958	55.6	69.1	74.8	66.52857	0.7	El Nino
11	1959	50.4	61.2	71.2	60.93626	1.2	El Nino
12	1960	52.5	65.2	73.2	63.65055	1	El Nino
13	1961	52.6	60	72.4	61.64835	0.8	El Nino
14	1962	53.8	64.8	70	62.88791	0.5	El Nino
15	1963	55.5	60.9	72.6	62.97692	0.5	El Nino
16	1964	48.1	58.4	70.4	58.96044	-1.4	La Nina
17	1965	53.9	60.3	71.7	61.94835	-0.6	La Nina
18	1966	53.3	65.2	70	62.85934	-0.6	La Nina
19	1967	48.4	58.9	72.4	59.88901	-0.6	La Nina
20	1968	49.7	65.6	71.3	62.23736	-0.8	La Nina
21	1969	49.6	59.4	72.1	60.35604	-0.8	La Nina
22	1970	53	58.9	72.7	61.5044	-0.5	La Nina
23	1971	54.2	59.9	68.9	60.98791	-0.9	La Nina
24	1972	47.1	64.3	69.7	60.40989	-0.9	La Nina
25	1973	55.6	61.5	74.1	63.70879	-0.5	La Nina

Fig. 2.7. Composite_XXX_MMMYYYY_cccc.xls showing the data sorted by episode.

	E	F	G	H	I	J	K	L	M
33	63.6022	-0.3	neutral		1999	53.4	63.1	74	63.4956
34	60.13516	0.7	El Nino		2000	51.1	63.4	72.3	62.27912
35	61.00989	1	El Nino						
36	61.72857	-0.5	La Nina		STATION COMPOSITES				
37	62.45714	-0.8	La Nina		El Nino	APR	MAY	JUNE	AMJ
38	63.35604	-0.2	neutral		Below	3	6	4	5
39	63.0033	1	El Nino		Near	6	4	4	2
40	61.70549	-0.9	La Nina		Above	5	4	6	7
41	62.26923	-0.6	La Nina		total				14
42	61.67802	0.2	neutral		La Nina	APR	MAY	JUNE	AMJ
43	66.52857	0.7	El Nino		Below	9	7	6	7
44	60.93626	1.2	El Nino		Near	2	5	6	7
45	63.65055	1	El Nino		Above	6	5	5	3
46	64.40879	0.4	neutral		total				17
47	62.26813	0.2	neutral		Neutral	APR	MAY	JUNE	AMJ
48	61.56044	-0.1	neutral		Below	7	9	7	8
49	61.64835	0.8	El Nino		Near	6	6	10	5
50	62.88791	0.5	El Nino		Above	10	8	6	10
51	63.4956	-0.7	La Nina		total				23

Fig. 2.8 The Station Composites table is shown highlighted by the orange box. For each event, El Niño, La Niña, Neutral, the number of events categorized by Below, Near, or Above normal are shown for each month and 3-month period with the total number of events.

d. Highlight the Neutral events.

i. Count the number of Neutral events in the below (blue), normal (black), and above normal (red) categories for each separate month and for the 3-month season. For example, in April for Neutral, there are 7 below, 6 near, and 10 above.

ii. Write these numbers in the appropriate cells J48 to M50 (Fig. 2.8). The total event column will automatically fill in as you enter data into the table.

iii. Make sure that the total number of events for each episode is the same for each month and 3-month season.

After completing steps 8b-d, you should have three 4x3 tables with numbers completed in the station composite table at this point (Fig. 2.8, orange square area).

e. Station Composite Analysis Probabilities. As you complete the steps above, you will see the table beginning at cell O38 automatically filling in (Fig. 2.9). These values represent the raw percentage for each event, e.g.,

0.357143 means during AMJ 35.7143% of the years during an El Niño have below normal temperatures.

Cell T38-V40 converts the station composite analysis probabilities to percent and cell Y38-AA40 rounds off the percent.

	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
36	STATION COMPOSITE PROBABILITIES													
37	AMJ	El Nino	Neutral	La Nina		AMJ	El Nino	Neutral	La Nina		AMJ	El Nino	Neutral	La Nina
38	Below	0.357143	0.347826	0.411765		Below	35.71428571	34.7826087	41.17647059		Below	35.7	34.8	41.2
39	Near	0.142857	0.217391	0.411765		Near	14.28571429	21.73913043	41.17647059		Near	14.3	21.7	41.2
40	Above	0.5	0.434783	0.176471		Above	50	43.47826087	17.64705882		Above	50.0	43.5	17.6

Fig. 2.9. Composite_XXX_MMMYYYY_cccc.xls showing the 3-month season raw probabilities of having below, near, above normal temperatures during an El Niño, Neutral, and La Niña based on historical composite analysis.

f. Re-sort the 1950-present data table from cell A2 through G54 based on column A to restore it to chronological order. From the *"Data"* menu, select *"Sort"*. Make sure *"no header row"* is selected, and then select column A to sort by.

g. Save the spreadsheets. You have completed the historical composite analysis for that climate variable for that station. Continue to complete more climate variables for this station, and for all your composite analysis stations.

h. Calculate the mean 3-month value of the climate variable.
i. In column E, under the last value, type:
        =Average(E2:Ex)
    where x is the number of the last cell with a 3-month average value.

ii. Format the mean average temperature. Highlight the mean average temperature. Right mouse click on the cell. Select *"Format Cells"*. Go to *"Number Tab"*. Under *"Category"* select *"Number"* and decimal of 1 place. Click *"OK"*.

iii. In the cell next to this, type "3-month average Temperature".

iv. Create a graphic for the historical composite analysis. In the template this is already formatted. You will need to edit the chart title and add the mean of the climate variable. Go to section 6 for detailed instructions on how to create a graphic manually.

Section 3: Trend Adjustment **Last revised 6.10.2005**

Trend Adjustment is not included here, but will be part of the operational procedure included in the algorithms being developed. There is no additional work on the part of the local office.

Trend adjustment is performed to correct the climate variable deviating from the expected trend. When there is an equal distribution of events among the 3 categories, above, near and below, the trend does not need to be adjusted. However, if there are more colder or neutral years than expected during a relatively warm period, the trend needs to be adjusted to reflect the actual observations. The trend adjustment is based on the previous 10 years of data, and if the deviation from the trend is statistically, then the trend adjustment is applied using a least squares fit method.

Section 4: Risk Analysis (Statistical Significance)

Last revised 6.10.2005

This section will assist you in understanding the risk analysis methodology and provide guidance on creating a formatted Excel spreadsheet from scratch.

Risk Analysis Background/Methodology

Risk analysis assesses whether the composite forecast is statistically significant. A hypergeometric distribution is used as a proxy to describe the probability distribution between all possible outcomes of a category within El Niño or La Niña years. A hypergeometric distribution is used in sampling without replacement from a finite population (each observation is either a success or a failure) and returns a probability that describes the probability of all possible outcomes.

The Microsoft Excel function, HYPGEOMDIST, will be used. This function is based on the following equation:

$$P(X=x)=h(x,n,M,N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \qquad (1)$$

From Microsoft Excel 2002 Help, HYPGEOMDIST

where:

Table 4.1 Explanation of hypergeometric variables

Variable (Excel variable)	Description	Example: For Above La Niña-see Fig. 4.1 & 4.2
x (sample_s)	the number of successes in the sample = the # above, near, **or** below for El Niño, Neutral, **or** La Niña	For Above La Niña, x=3 (there were 3 Above La Niña events counted in the dataset). This is used to determine that statistical significance. If the # of events falls within either the upper or lower tail of the hypergeometric distribution, then it is statistically significant.
n (number_sample)	the size of the sample =	n = 17

	# of El Niño, Neutral, or La Niña	total number of La Niña events
M (population_s)	the number of successes in the population = total # above, near, **or** below for El Niño, Neutral **and** La Niña	M = 20 sum of number of above events for La Niña, Neutral, and El Niño
N (number_population)	the population size = total # El Niño, Neutral, La Niña events	N = 54 total # of events

and:

=HYPGEOMDIST (x,n,M,N)

with x, n, M, N corresponding to the appropriate cells in the spreadsheet (Fig. 4.2).

The value returned will be a probability (decimal). There are some instances where #NUM! error value will appear, these include:

If $x < 0$ or $x >$ the lesser of N or M,
If $x <$ the larger of 0 or (n-N+M),
If $n < 0$ or $n > N$,
If $M < 0$ or $M > N$, or
If $N < 0$

	E	F	G	H	I	J	K	L	M
33	63.6022	-0.3	neutral		1999	53.4	63.1	74	63.4956
34	60.13516	0.7	El Nino		2000	51.1	63.4	72.3	62.27912
35	61.00989	1	El Nino						
36	61.72857	-0.5	La Nina		STATION COMPOSITES				
37	62.45714	-0.8	La Nina		El Nino	APR	MAY	JUNE	AMJ
38	63.35604	-0.2	neutral		Below	3	6	4	5
39	63.0033	1	El Nino		Near	6	4	4	2
40	61.70549	-0.9	La Nina		Above	5	4	6	7
41	62.26923	-0.6	La Nina		total				14
42	61.67802	0.2	neutral		La Nina	APR	MAY	JUNE	AMJ
43	66.52857	0.7	El Nino		Below	9	7	6	7
44	60.93626	1.2	El Nino		Near	2	5	6	7
45	63.65055	1	El Nino		Above	6	5	5	3
46	64.40879	0.4	neutral		total				17
47	62.26813	0.2	neutral		Neutral	APR	MAY	JUNE	AMJ
48	61.56044	-0.1	neutral		Below	7	9	7	8
49	61.64835	0.8	El Nino		Near	6	6	10	5
50	62.88791	0.5	El Nino		Above	10	8	6	10
51	63.4956	-0.7	La Nina		total				23

Fig. 4.1 For the Above La Niña case, the green pentagon represents # of events, the blue arrow is # of all La Niña cases – n, the sum of the pink squares represents M, and the sum of the black circles represents the total # of cases N. These values then populate the corresponding location in the Risk Analysis table shown in Fig. 4.2.

	H	I	J	K	L	M	N	O
53	RISK ANALYSIS		Above	Above	Above	Below	Below	Below
54			La Nina	Neutral	ElNino	La Nina	Neutral	ElNino
55		n	17	23	14	17	23	14
56		x	3	10	7	7	8	5
57		M	20	20	20	20	20	20
58		N	54	54	54	54	54	54
59		outcome # ranges from 0 to n						

Fig. 4.2 The Risk Analysis Table shows each event (e.g., Above La Niña) and the associated value of the values used for calculations in the hypergeometric distribution.

Risk Analysis Procedure/Instructions

Using the Template
Using the Composite_XXX_MMMYYYY_cccc.xls spreadsheet, the risk analysis will automatically fill in after section 2: Compute Historical Composites Analysis is completed, **BUT you must check some parts of the setup manually**. (skip to *Note 2* below if using the template)

Setting up a Template from Scratch
Note 1: The instructions below ***must be*** performed in the stated cells in the Excel worksheet.

a. Set up the "Risk Analysis Table" exactly (your cells should look the same) as shown in Fig. 4.3. This table will contain the variables needed to calculate the probabilities to determine whether the category and event is statistically significant from equation 1. Format each cell from the information in Fig. 4.1, in the Station Composite section. Table 2 provides the information that needs to be input into the corresponding cells shown in Fig. 4.3 in the spreadsheet you are creating from scratch.

	H	I	J	K	L	M	N	O
53	RISK ANALYSIS		Above	Above	Above	Below	Below	Below
54			La Nina	Neutral	ElNino	La Nina	Neutral	ElNino
55		n						
56		x						
57		M						
58		N						
59		outcome # ranges from 0 to n						

Fig. 4.3 Risk Analysis Table set up

Table 2. Formatting information for Fig. 4.3

	J	K	L	M	N	O
	Above La Nina	Above Neutral	Above El Nino	Below La Nina	Below Neutral	Below El Nino
n	=M46	=M51	=M41	=M46	=M51	=M41
x	=M45	=M50	=M40	=M43	=M48	=M38
M	=M40+M45+M50	=M40+M45+M50	=M40+M45+M50	=M38+M43+M48	=M38+M43+M48	=M38+M43+M48
N	=M41+M46+M51	=M41+M46+M51	=M41+M46+M51	=M41+M46+M51	=M41+M46+M51	=M41+M46+M51

b. Setup a table exactly (your cells should look the same) as in Fig. 4.4. ***for each***: Above La Niña, Above Neutral, Above El Niño, Below La Niña, Below Neutral, and Below El Niño.

	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
62	Above	Outcome	0	1	2	3	4	5	6	7	8	9	10	11	12
63	La Nina	P(x)													
64		Sum P(x)													
65		1-Sum P(x)													

Fig. 4.4 Risk analysis table before formatting for the Above La Niña outcome only.

c. **Manual** **formatting instructions for P(x), Sum P(x), and 1- Sum P(x).**
For the parameters: P(x), Sum P(x), and 1 – Sum P(x), each cell must be
formatted with the correct formula to calculate each parameter.

The following Table 3 shows the equations to format the outcome=0 cells to
calculate P(x), Sum P(x), and 1 – Sum P(x). Only outcome =0, 1, 2 or shown, but
this would continue in the spreadsheet until the outcome number = n, following
this format. Use this as a guide for the others (above neutral, above El Niño,
below La Niña, below neutral, and below El Niño).

Table 3 Formatting guide to calculate P(x), Sum P(x), and 1 – Sum P(x)

Outcome	0	1	2
P(x)	In cell J63, type: =HYPGEOMDIST(J62,J55,J57,J58) where J62 = x, J55 = n, J57 = M, J58 = N	In cell K63, type: =HYPGEOMDIST(K62,J55,J57,J58)	In cell K64, type: =HYPGEOMDIST(L63, J55, J57, J58)
Sum P(x)	In cell J64, type: =J63	In cell K64, type: =J64+K63	In cell L64, type: = K64+L63
1 – Sum P(x)	In cell J65, type: =1	In cell K65, type: =1-J64	In cell L65, type: =1-K64

Note 2: The Outcome number will vary for each (Above La Niña, Above Neutral,
etc.) depending on the n value. The outcome number is the number of possible
outcomes of a specific event (e.g., Above La Niña), which ranges from 0 to n. For
Above La Niña, the number of possible outcomes ranges from 0 to 17, since 17
Above La Niña events were counted from the dataset. You should have 0…17 in
the Outcome row (see Fig. 4.5 – note the figure is cut off before outcome =17).

On the template, this has been done, **BUT** depending on the value of n you may
need to add additional formatted cells. ***This must be determined and entered***
manually. This only needs to be completed if outcome is greater than 25.

c. Format information for Fig. 4.4:

If n>25, continue with the information below; otherwise go the next step (d).

For the parameters: Num, P(x), Sum P(x), and 1 – Sum P(x), each cell must be
formatted with the correct formula to calculate each parameter.

The following Table 4 shows the equations to format the outcome=0 cells to
calculate P(x), Sum P(x), and 1 – Sum P(x). Only outcome =0, 1, 2 or shown, but
this would continue in the spreadsheet until the outcome number = n, following

this format. Use this as a guide for the others (above neutral, above El Niño, below La Niña, below neutral, and below El Niño *if n>25*).

Table 4. Formatting guide to calculate P(x), Sum P(x), and 1 – Sum P(x)

Outcome	26	27 etc.
P(x)	In cell AJ63, type: =HYPGEOMDIST(AJ62,J55,J57,J58) where AJ62 = x, J55 = n, J57 = M, J58 = N	In cell AK63, type: =HYPGEOMDIST(AK62,J55,J57,J58)
Sum P(x)	In cell AJ64, type: =AJ63 + AI64	In cell AK64, type: = AK63 + AJ64
1 – Sum P(x)	In cell AJ65, type: =1 - AI64	In cell AK65, type: =1 - AJ64

d. Format the upper and lower tails of the hypergeometric distribution. *This must be done <u>manually</u> for the hypergeometric distributions values.*

i. Format the following: P(x), Sum P(x), and 1 – Sum P(x) for each Above La Niña, Above Neutral, Above El Niño, Below La Niña, Below Neutral, Below El Niño using the instructions in Table 5. Information is also provided about what each of these values represents. The results can be seen in Fig. 4.5.

**Table 5.** Color formatting instructions for P(x) and distribution tails

| What to format	How to format	What the formatted values represent
**P(x) = x	**Black bold** and <u>underline</u>	represents the value that must be compared against the tails of the distribution to determine statistical significance
Sum P(x) ≤ 0.10	**Blue bold**	represents the left tail of the distribution
1 – Sum P(x) ≤ 0.10	**Red bold**	represents the right tail of the distribution

**See Fig. 4.2 or 4.3 to see how to locate the x value.

F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
Above	Outcome	0	1	2	3	4	5	6	7	8	9	10	11	
La Nina	P(x)	4.94897E-05	0.000934806	0.007478	**0.033653**	0.09535	0.180299	0.235172	0.215574	0.140123	0.064672	0.021078	0.004791	0.
	Sum P(x)	**0.000**	**0.001**	**0.008**	**0.042**	0.137	0.318	0.553	0.769	0.909	0.973	0.994	0.999	
	1-Sum P(x)	1	1.000	0.999	0.992	0.958	0.863	0.682	0.447	0.231	**0.091**	**0.027**	**0.006**	
Above	Outcome	0	1	2	3	4	5	6	7	8	9	10	11	
Neutral	P(x)	2.63459E-07	1.00993E-05	0.000162	0.001461	0.008281	0.031466	0.083293	0.157332	0.215296	0.215296	**0.157884**	0.084814	0.0
	Sum P(x)	**2.63459E-07**	**1.03627E-05**	**0.000173**	**0.001634**	**0.009915**	**0.041381**	0.124674	0.282006	0.497302	0.712597	0.870481	0.955295	0.
	1-Sum P(x)	1	0.999999737	0.99999	0.999827	0.998366	0.990085	0.958619	0.875326	0.717994	0.502698	0.287403	0.129519	**0.0**
Above	Outcome	0	1	2	3	4	5	6	7	8	9	10	11	
El Nino	P(x)	0.000428911	0.005718811	0.032103	0.100497	0.195761	0.250573	0.216842	**0.128499**	0.052203	0.014401	0.00264	0.00031	8.1
	Sum P(x)	**0.000428911**	**0.006147722**	**0.038251**	0.138748	0.334509	0.585082	0.801925	0.930424	0.982627	0.997028	0.999668	0.999977	0.9
	1-Sum P(x)	1	0.999571089	0.993852	0.961749	0.861252	0.665491	0.414918	0.198075	**0.069576**	**0.017373**	**0.002972**	**0.000332**	2.2
Below	Outcome	0	1	2	3	4	5	6	7	8	9	10	11	
La Nina	P(x)	4.94897E-05	0.000934806	0.007478	0.033653	0.09535	0.180299	0.235172	**0.215574**	0.140123	0.064672	0.021078	0.004791	0.
	Sum P(x)	**4.94897E-05**	**0.000984295**	**0.008463**	**0.042116**	0.137466	0.317764	0.552936	0.768511	0.908634	0.973306	0.994385	0.999175	0.9
	1-Sum P(x)	1	0.99995051	0.999016	0.991537	0.957884	0.862534	0.682236	0.447064	0.231489	**0.091366**	**0.026694**	**0.005615**	**0.0**
Below	Outcome	0	1	2	3	4	5	6	7	8	9	10	11	
Neutral	P(x)	2.63459E-07	1.00993E-05	0.000162	0.001461	0.008281	0.031466	0.083293	0.157332	**0.215296**	0.215296	0.157884	0.084814	0.0
	Sum P(x)	**2.63459E-07**	**1.03627E-05**	**0.000173**	**0.001634**	**0.009915**	**0.041381**	0.124674	0.282006	0.497302	0.712597	0.870481	0.955295	0.
	1-Sum P(x)	1	0.999999737	0.99999	0.999827	0.998366	0.990085	0.958619	0.875326	0.717994	0.502698	0.287403	0.129519	**0.0**
Below	Outcome	0	1	2	3	4	5	6	7	8	9	10	11	
El Nino	P(x)	0.000428911	0.005718811	0.032103	0.100497	0.195761	**0.250573**	0.216842	0.128499	0.052203	0.014401	0.00264	0.00031	2.1
	Sum P(x)	**0.000428911**	**0.006147722**	**0.038251**	0.138748	0.334509	0.585082	0.801925	0.930424	0.982627	0.997028	0.999668	0.999977	1.0
	1-Sum P(x)	1	0.999571089	0.993852	0.961749	0.861252	0.665491	0.414918	0.198075	**0.069576**	**0.017373**	**0.002972**	**0.000332**	2.2

Fig. 4.5 Risk Analysis data showing the probability P(x) for each outcome. Blue (red) values represent the left (right) tail of the distribution, and the bold underlined value is the probability for the outcome to be compared to the left and right tails to determine statistical significance.

The probability distribution for the above temperature category during La Niña is shown in Fig. 4.6 and 4.7. Figure 4.6 shows the left tail of the distribution in blue and Fig. 4.7 shows the right tail of the distribution in red. Both tails indicate a statistical significance level of 10% (90% confidence).
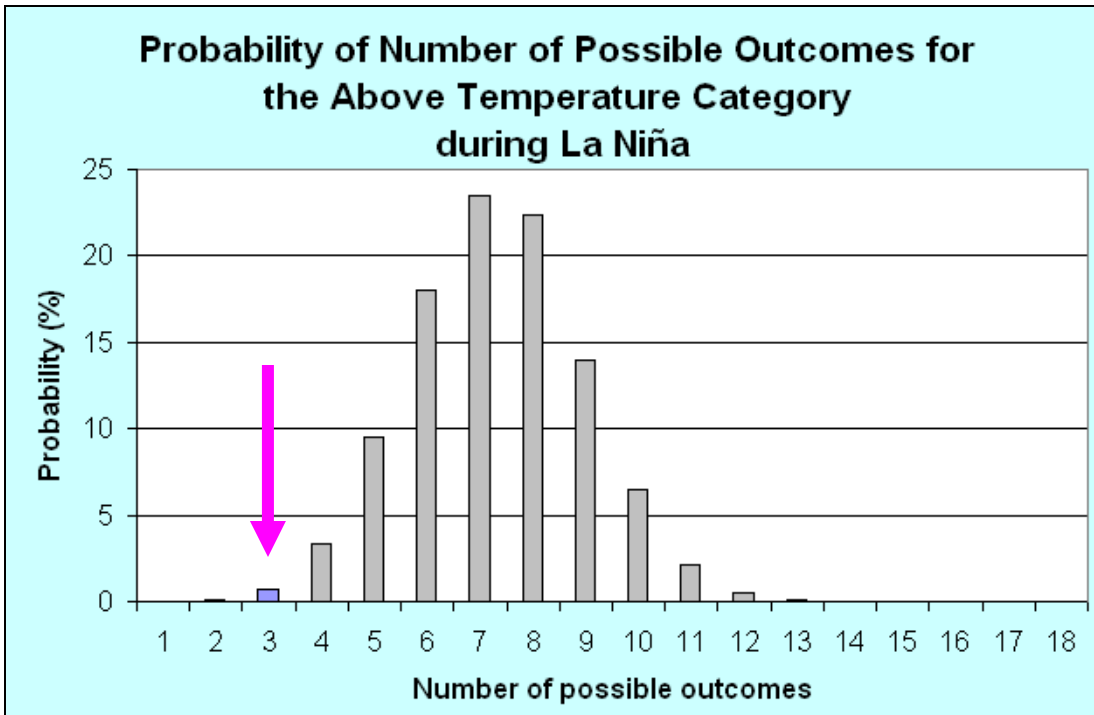
Fig. 4.6. The probability distribution for all possible outcomes for the above temperature category during La Niña. The right tail of the distribution is shown in blue. The pink arrow points to P(x)=3 (this is the number of actual observed above temperature events during La Niña) and is within the right tail, so this is statistically significant.
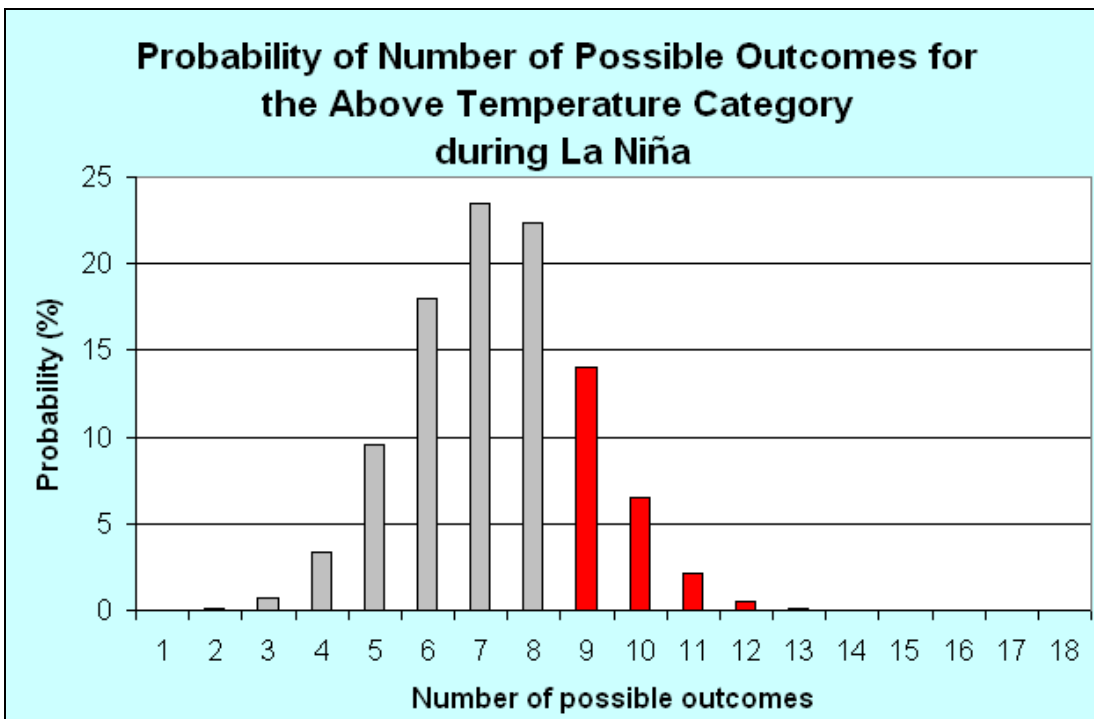
Fig. 4.7. The probability distribution for all possible outcomes for the above temperature category during La Niña. The left tail of the distribution is shown in red.

ii. Determining Statistical Significance

A 10% statistical significance level will be used (this is the same as a 90% confidence interval). The values formatted in blue and red represent the 10% lower (blue) and upper (red) "tails" of the probability distribution function (bell curve) and define the values for "significance". The statistical significance needs to be manually assessed, this is not automated.

Determining Statistical Significance Test

Is the black bold value in the spreadsheet within the upper (red) <u>or</u> lower (blue) tail?

   If Yes, then it is statistically significant!

   If No, then it is <u>NOT</u> statistically significant.

For example, in Fig. 4.5, for Above La Niña, x=3 and the bold value, P(x) = 0.033653, falls within the blue values, 0 – 0.042. **Therefore, if the black, bold value is within the tail (either within the red or blue values), the historical composite analysis is statistically significant to the 10% (p = 0.10) level.**

If ***any*** of the risk analysis values (above La Niña, above Neutral, above El Niño, below La Niña, below Neutral, below El Niño) ***show*** statistical significance, then proceed to **making the composite analysis forecast**.

If ***none*** of the risk analysis values (above La Niña, above Neutral, above El Niño, below La Niña, below Neutral, below El Niño) ***does not*** show statistical significance, then **no forecast is made** and instead climatology will be provided.

e. Indicate statistical significance on the historical composite analysis graphic. Statistical significance is indicated in Fig. 4.8 by the bold line around the bar corresponding to the above temperature category during La Niña. Section 6 provides more instructions on how to make a bold line around a temperature bar.
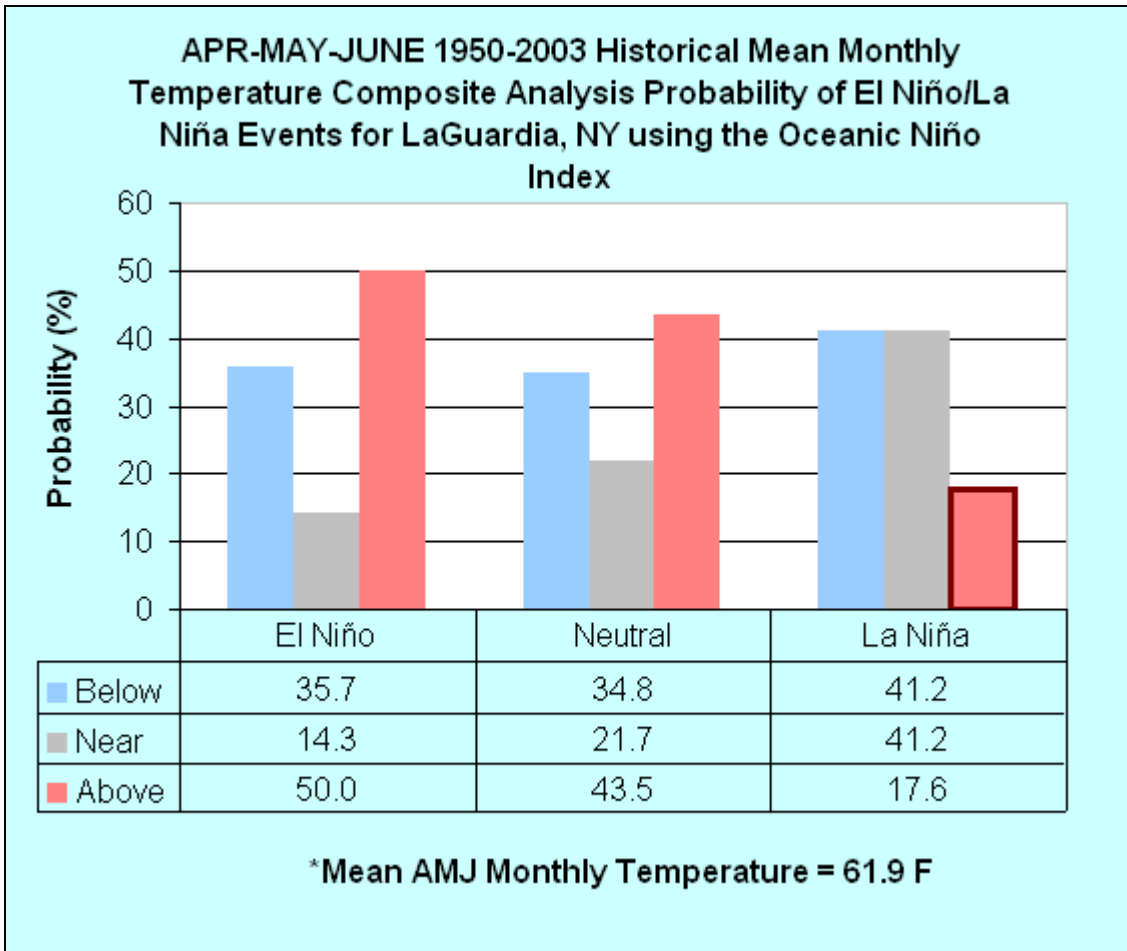
Fig. 4.8. Historical composite analysis with statistical significance indicated by heavy black line around above La Niña.

Section 5: Composite analysis forecast

The Composite Analysis Forecast

1.   Make the composite analysis forecast.

   a. Copy the 3-month period CPC Forecast (consolidated forecast for Niño 3.4 SST) from
   http://www.cpc.ncep.noaa.gov/products/precip/CWlink/ENSO/sstcon34.txt
   into cell P45 to R45 (Fig. 2.10).

   b. Make the forecast. *This is already formatted in the template, below is an explanation of how the forecast is calculated.*

   If using the template, the forecast will automatically fill-in cell P48 to R48 as previous steps are completed (Fig. 5.1).

   For the 3-month period, the equations to compute the forecast probabilities are as follows:

$$P_{above}^{station} = P_{above/ElNino}^{station} * P_{above}^{Nino3.4} + P_{above/Neutral}^{station} * P_{near}^{Nino3.4} + P_{above/LaNina}^{station} * P_{below}^{Nino3.4} \quad (1)$$

$$P_{near}^{station} = P_{near/ElNino}^{station} * P_{above}^{Nino3.4} + P_{near/Neutral}^{station} * P_{near}^{Nino3.4} + P_{near/LaNina}^{station} * P_{below}^{Nino3.4} \quad (2)$$

$$P_{below}^{station} = P_{below/ElNino}^{station} * P_{above}^{Nino3.4} + P_{below/Neutral}^{station} * P_{near}^{Nino3.4} + P_{below/LaNina}^{station} * P_{below}^{Nino3.4} \quad (3)$$

   This translates into the following: the above composite analysis forecast (P_{above}) is calculated by taking "the probability to be in the above normal category given El Niño" x the "forecast probability of Niño 3.4 in the above category" + the "probability to be in the above normal category given Neutral" x the "forecast probability of Niño 3.4 in the near normal category" + the "probability to be in the above normal category given La Niña" x the "forecast probability of Niño 3.4 in the below normal category", or

   P_{above} = (0.5 x 0.176) + (0.435 x 0.806) + (0.176 x 0.017)   (4)

Equation (4) uses data from Fig. 2.9 in section2 cells O40 to Q40 and Fig. 5.1 cells P45 to R45.

Similarly, for P_{near} and P_{below}:

$$P_{near} = (0.143 \times 0.176) + (0.217 \times 0.806) + (0.412 \times 0.017) \qquad (5)$$

$$P_{below} = (0.357 \times 0.176) + (0.348 \times 0.806) + (0.412 \times 0.017) \qquad (6)$$

The above forecast needs to be completed for each 3-month period (13 leads).

Note: Cells P50 to R50 convert the raw forecast to percent and cells P52 to R52 round of the percent to 1 decimal place.

	O	P	Q	R	S	T	U	V	W	X	Y
43	CPC Forecast (consolidated forecast for Nino 3.4 SST) http://www.cpc.ncep.noaa.gov/products/precip/CWlink/ENSO/sstcon34.txt										
44		Below	Normal	Above							
45	AMJ	0.017	0.806	0.176							
46	FORECAST based on station composites and CPC Nino 3.4 SST forecast										
47		Below	Normal	Above							
48	AMJ	0.350205	0.20736	0.441435							
49		Below	Normal	Above							
50	AMJ	35.0205	20.73602	44.14348		EC = Equal chance = 33%					
51		Below	Normal	Above	EC						
52	AMJ	35.0	20.7	44.1	33						

Figure 5.1. Composite_XXX_MMMYYYY_cccc.xls showing the composite analysis forecast.

Note: Equal changes is 33% for each category (below, near, above).

Section 6: Creating Composite Analysis Products

Last revised 6.10.2005

Creating Graphical Products

Two graphical products will be produced, a Historical Composite Analysis and a Composite Analysis Forecast, for each 3-month period. *The exact product presentation has not been fully developed yet. The final operational design may be different.*

1. Historical Composite Analysis Graphic (Fig. 6.1).
Note: In the template, this is already formatted!
You must edit the chart title and add the mean of the climate variable.

Creating a historical composite analysis graphic <u>manually:</u>

a. Select cells Y37 to AA40 and click on the chart wizard icon.

b. Select *"Column"* as the type of chart. Click *"Next"*.

c. On the *"Data Range"* tab, under *"Series In"* select *"rows"*. Click *"Next"*.

d. On the *"Title"* tab, type a label for the x and y axes and type a descriptive chart title. On the *"Data Table"* tab, check *"show data table"*. Click *"Next*

e. Select *"Save as an object in"*. Click *"Finish"*.

f. Change/modify the chart as needed.

g. Add the EC line to the chart (EC is equal chances = 33%).

h. Add a text box below the chart and add the mean climate parameter value for the 3-month period.

Indicating Statistical Significance on the Historical Composite Analysis graphic (Fig. 6.1)
Statistical significance can be shown on a historical composite analysis graphic. In Fig. 6.1, statistical significance is indicated by a bold outline around the bar on the graph for Above La Nina.
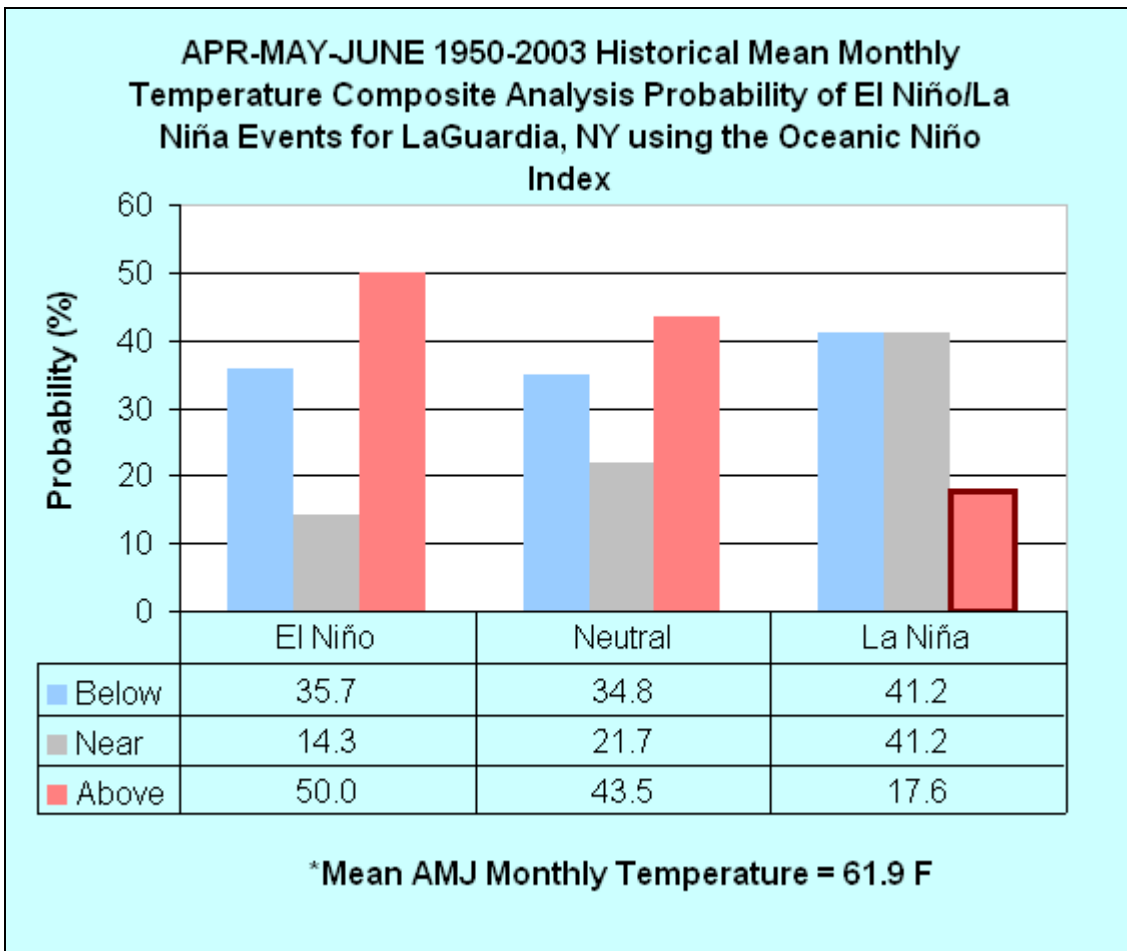
To create a bold line around a bar:

From the Risk Analysis section, is any of the data statistically significant?

> **Yes,** continue with instructions below to place a bold line around the appropriate bar for the event in the figure you created.

1. Right mouse click on the bar and a square will appear in the center of the bar.

2. Left mouse click inside the bar. The bar will appear selected.

3. Right mouse click inside the bar. A menu will appear. Select *"Format Data Point"*.

4. On the *"Patterns"* tab, under *"Border"* select *"Custom"* and select the color and weight (of the line to be place around the bar).

5. Click *"OK"*. A bold line should now be around the bar.

APR-MAY-JUNE 1950-2003 Historical Mean Monthly Temperature Composite Analysis Probability of El Niño/La Niña Events for LaGuardia, NY using the Oceanic Niño Index

	El Niño	Neutral	La Niña
Below	35.7	34.8	41.2
Near	14.3	21.7	41.2
Above	50.0	43.5	17.6

*Mean AMJ Monthly Temperature = 61.9 F

Fig. 6.1. Historical composite analysis probability based on El Nino/La Nina. Statistical significance is indicated by heavy black border around the bar for the above temperature category for La Nina.
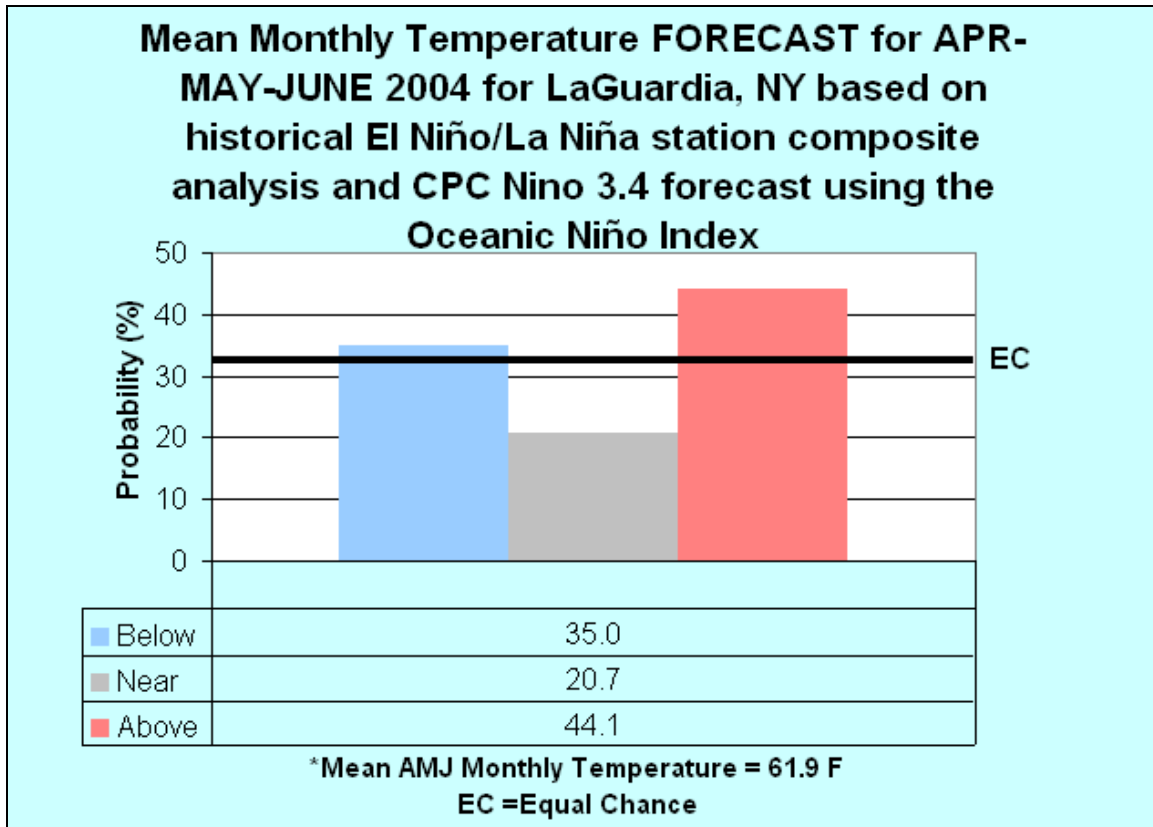
2.   Composite Analysis Forecast Graphic.
Note: In the template, this is already formatted! You must edit the chart title and add the mean climate parameter value.

Creating a composite analysis forecast graphic <u>manually:</u>

a. Bar Graph (Fig 6.2).

>   i. Select cells P51 to R52 and click on the chart wizard.
>
>   iI. Select *"Column"* as the type of chart. Click *"Next".*
>
>   iii. On the *"Data Range"* tab, under *"Series In"* select *"columns"*. Click *"Next".*
>
>   iv. On the *"Title"* tab, type a label for the x and y axes and type a descriptive chart title. On the *"Data Table"* tab, check *"show data table".* Click *"Next".*
>
>   v. Select *"Save as an object in"*. Click *"Finish".*
>
>   vi. Change/modify the chart as needed.
>
>   vii. Add the EC line to the chart (EC is equal chances = 33%) by drawing a line and adding 'EC' in a text box (see Fig. 6.2).
>
>   viii. Add a text box below the chart and add the mean climate parameter value for the 3-month period.
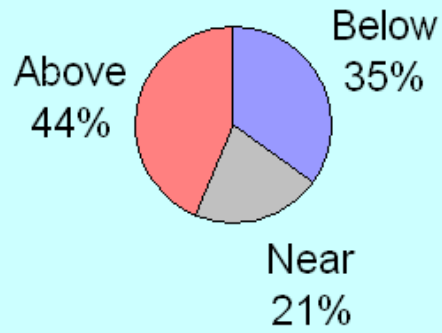
Fig. 6.2 Composite analysis forecast for a 3-month period. Equal chance (EC; 33%) is indicated by the heavy black line.

b. Pie Chart Graphic (Fig. 6.3).

i. Select cell P51 to R52.

ii. Click on chart wizard and select *"pie"* as the chart type chart. Click *"Next"*.

iii. On the *"Data Range"* tab, select *"rows"*. Click *"Next"*.

iv. On the "Title" tab, type a label for the x and y axes and type a descriptive chart title. On the "Data Labels" tab, check *"category name"* and *"percentage"*. Click *"Next"*.

v. Select *"Save as an object in"*. Click *"Finish"*.

vi. Change/modify the chart as needed.

vi. Add a text box below the pie chart and add the mean climate parameter value for the 3-month period.

Mean Monthly Temperature FORECAST for APR-MAY-JUNE 2004 for LaGuardia, NY based on historical El Niño/La Niña station composite analysis and CPC Nino 3.4 forecast using the Oceanic Niño Index

Above 44%

Below 35%

Near 21%

* Mean AMJ Monthly Temperature = 61.9 F

Fig. 6.3. Composite analysis forecast for a 3-month period. Equal chance is 33%.

Section 7: Hindcast and Forecast Verification

Last revised 6.10.2005

Hindcast Forecast

Hindcast Forecast is not included here, but will be part of the operational procedure included in the algorithms being developed. The local office is responsible for interpretation of the hindcast forecast.

Verification is usually done on past forecasts; however, since composite analysis is a new forecasting tool, past forecasts are not currently available. Therefore, verification will use hindcast forecasts of station composite analysis and archived CPC Nino 3.4 forecasts. CPC El Niño/La Niña forecasts are available from 1982 to the present. The hindcast forecast will be compared to the observations, and different measures of verification, like Ranked Probability Score, Ranked Probability Skill Score, and the Heidke Skill Score will be produced.

Forecast Verification

Composite Analysis Forecast Verification is also not included here, but will be part of the operational procedure included in the algorithms being developed. The local office is responsible for interpreting the verification results.

Verification provides information about the quality, reliability, and confidence of the product. This information can then serve as guidance for improving the product and help customers with their decision-making. Verification on local composite analysis forecast products will be centrally produced and provided to field offices.

Section 8: Resources

Web pages

CPC Niño 3.4 SST consolidated forecast

http://www.cpc.ncep.noaa.gov/products/precip/CWlink/ENSO/sstcon34.txt

Seasonal ONI (1950-present)

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

Seasonal Climate Prediction for Regional Scales

http://www.ccnmtl.columbia.edu/projects/climate/

xmACIS – Applied Climate Information System

http://xmacis.rcc-acis.org

Climate Professional Development Series, Instructional Unit 4

http://www.nws.noaa.gov/om/csd/pds/pcu4/index.htm

Climate Operations Course

http://www.comet.ucar.edu/class/common/html/course_descrip.html#

Literature

Barnston, A. G., A. Leetmaa, V. E. Kousky, R.E. Livezey, E. A. O'Lenic, H. Van den Dool, A. J. Wagner, and D. A. Unger, 1999: NCEP Forecasts of the El Niño of 1997–98 and Its U.S. Impacts., *Bull. Amer. Met. Soc.*, **80**, 1829–1852.

Montroy, D.L., M. B. Richman, and P. J. Lamb, 1998: Observed Nonlinearities of Monthly Teleconnections between Tropical Pacific Sea Surface Temperature Anomalies and Central and Eastern North American Precipitation**, *J. Climate*, **11**,1812–1835.

Wilks, D.S., 1995: *Statistical Methods id Atmospheric Sciences: An Introduction*. Chapter 7. Forecast Verification. Acadmeic Press, 233–283 pp.