Crop Monitoring as an E-agricultural tool in Developing Countries

STATISTICAL TOOLBOX PILOTING AND WORKSHOP REPORT

Held at Kenitra from February 15 to 17, 2012

Reference: ***E-AGRI_D62.2_PilotingWorkshopReport***
Author(s): Steven Hoek
Version: 2.0
Date: 21/02/2013

DOCUMENT CONTROL

Signatures

Author(s) : Steven Hoek

Reviewer(s) : Allard de Wit
Qinghan Dong

Approver(s) :

Issuing authority : Alterra

Change record

Release	Date	Pages	Description	Editor(s)/Reviewer(s)
1.0	06/03/2012	3	First draft of deliverable. E-Agri document template not yet used	Steven Hoek
2.0	20/02/2013	12	E-Agri document template used	Steven Hoek
2.1	21/02/2013	44	Bundled exercises added as Annex 2	Steven Hoek

TABLE OF CONTENT

1. Introduction 7
1.1. Day One 7
1.2. Day Two 8
1.3. Day Three 9
2. Lessons learnt 10
3. Epilogue 11
Annex 1. List of Participants 12
Annex 2. Exercises 13
A. Exercises with the MARS Viewer 13
B. Training course on the CGMS Statistical Toolbox 26

ACRONYMS & GLOSSARY

Abbreviation	Description
AVHRR	Advanced Very High Resolution Radiometer (sensor)
CGMS	Crop Growth Monitoring System (model software)
CRRA	Centre Régional de Recherche Agronomique
CST	CGMS Statistical Tool (model software)
DMN	Direction de la Météorologie Nationale (organisation)
DMP	Dry Matter Production (indicator)
DSS	Direction de la Stratégie et des Statistiques (organisation)
E-Agri	an FP7 research project - funded by the EU - designed to support the uptake of European ICT research results in developing economies
FTP	File Transfer Protocol
INRA	Institut national de la recherche agronomique (organisation)
JRC	Joint Research Centre (organisation)
MARS Unit	Monitoring Agricultural Resources Unit (organisation)
NDVI	Normalised Difference Vegetation Index (indicator)
NOAA	National Oceanic and Atmospheric Administration (organisation)
NUTS	Nomenclature des Unités Territoriales Statistiques
ODBC	Open Database Connectivity
PCA	Principal Component Analysis
URACRP	Unité de Recherche sur l'Amélioration et Conservation des Ressources Phytogénétiques (organisation)
VITO	Flemish Institute for Technological Research (organisation)
WOFOST	World Food Studies (model software)

EXECUTIVE SUMMARY

This document gives an account of the training session organised for a group of 8 people at Kenitra, Morocco in February 2012. Aim was to allow the participants to become familiar with two pieces of software provided in the framework of the E-Agri project. These pieces of software are relevant for crop yield forecasting in Morocco.

1. Introduction

The picture below shows all the participants as well as the undersigned. A list with their names is provided in Annex 1. We were received very nicely by Mr Hassan Benaouda (shown on the right), manager of the "Centre Regional de la Recherche Agronomique" (CRRA) at Kenitra, Morocco. The venue was the building of the "Unité de Recherche sur l'Amélioration et Conservation des Ressources Phytogénétiques" (URACRP). During the three days, Mr Benaouda took very good care of us as far as food and drinks were concerned.

Figure 1. Participants of the workshop in front of the venue building

1.1. Day One

On this day, the web viewer was introduced. It is a web application hosted on a web server in Wageningen: http://www.marsop.info/glw/bin/webviewer.html. Each participant was

given a username and password. In all, the participants were asked to do 7 exercises with the web viewer, which were all meant to make the participants familiar with the available data and the various options to retrieve, visualise and analyse the weather indicators in the form of maps and graphs and the effect of the weather on the crop indicators. All exercises pertained to Morocco and during the day the focus gradually from the technological possibilities to the relevance of the web viewer for crop yield forecasting in Morocco. In the last exercise, the bulletin of 2011 was discussed. The presentation with the exercises can be found on the E-Agri FTP site under folder /Morocco/Training_Feb_2012/Day1.

1.2. Day Two

On this day, the CGMS Statistical Tool or CST was dealt with. Some of the participants had difficulties with installing the tool on their computers. Those difficulties were mainly issues related to the fact that Windows is now 64-bits whereas the CST is still a 32-bits programme. In all, the participants were asked to do 3 exercises with the CST which were meant to enable the participants to carry out time trend analysis, to analyse data availability, to detect anomalies in the data, to carry out regression analysis, to select the best model and to arrive at a forecast based on that model.

Figure 2: Impression of the workshop

In addition, there was an exercise that was meant to enable them to carry out scenario analysis. This involves a principal components analysis (PCA) on the indicator data. On the basis of the results of the PCA, a number of similar years are identified and the realised yields for those years are then used as a basis for a forecast for the current year. The presentation with the exercises can be found on the E-Agri FTP site under the following folder: /Morocco/Training_Feb_2012/Day2.

1.3. Day Three

On day 3, we continued with the CST. The participants were asked to finish their work on the exercises. An example was included that demonstrated how yield forecasts can develop over time (regression and scenario analysis). This was very useful for our understanding. It showed that during the cropping season, the accuracy of the regression forecast increases until a certain date. After that date, the accuracy of the forecast did not improve further. This can be considered as a general pattern. For the scenario analysis – as it has been implemented in the CST – the accuracy of the forecast did not improve over time.

In addition, we had a look at the software that is available for running the WOFOST crop model – i.e. the WOFOST Control Centre. It was explained that with WOFOST a simulation is done for a specific point location. With the CGMS programme, in effect the same WOFOST model is applied but then for grid areas. In order to obtain reliable results with CGMS, WOFOST will have to be used first to calibrate the model for various locations across the country. The files needed for carrying out the exercises can be found on the E-Agri FTP site under the folder /Morocco/Training_Feb_2012/Day3.

Figure 3: Logo of the E-Agri project

2. Lessons learnt

During the training we had a lot of fruitful discussions. The following things were learnt:

- Within Morocco, the meaning of the administrative units in NUTS Level 1 is very limited. Before Mr Riad Balaghi already proposed to group the provinces into agrozones; within each of these zones more or less similar conditions can be found
- Difference between map of long-term rainfall as obtained from webviewer and map shown by Tarik El-Hairech. Rainfall sums from the viewer are a lot higher than those observed at the meteorological stations, even when the crop mask is applied.
- The bulletin of last year was used in one of the exercises. It was produced by Riad Balaghi and Mohamed El Aydam. It included a paragraph stating that a longer vegetative cycle would affect yields favourably. This is the case when rainfall is high, as was the case during the time the bulletin was released. Otherwise, radiation in Morocco is always enough. Rainfall and heat after March are rather the phenomena posing constraints to yield development.
- So far, the CST has not been tested sufficiently. In fact, it has been used by Alterra and JRC only to a limited extent so far, with only few indicators. In fact, the participants did a good job in testing the software: several bugs were found in CST. These will be fixed in the foreseeable future.
- Major installation problems were encountered by those having Microsoft Office 64-bits installed on their computers. Conclusion is that there is the need for a 64-bits version of the CST for those users.
- Even for those with Microsoft Office 32-bits installed, it seemed to be impossible to add an extra data source – also known as ODBC link – in the ODBC Data Source Administrator. However, this was sorted out and solved by Alterra in Wageningen.
- Yields obtained in 2011 in the various provinces were not yet added to the database, because they were made available at a late point in time. Remote sensing indicators (NDVI and DMP) were highly correlated to official yields at both national and NUTS levels. However, the crop indicators were not correlated to cereal yields for Morocco, most likely because they are based on sowing dates valid for Europe. Hence, there is a need for WOFOST calibration.

3. Epilogue

Special thanks to Mr Riad Balaghi for organising the training. We look forward to cooperating further. Since the return of the undersigned to Wageningen, one the most important bugs in the CST has been fixed. Moreover, the NUTS level 1 in the database of the CST has been replaced by the AgroZone level. Crop and remote sensing indicator data have been made available for this level as well as for the provincial level. NDVI data based on NOAA / AVHRR – available for the years since 1982! - were also made available in the database. More bugs of the CST will be fixed in the foreseeable future.

Annex 1. List of Participants

Name	Affiliation	E-mail address
Balaghi, Dr Riad	INRA, Rabat	riad.balaghi@gmail.com
Benaouda, Mr Hassan	URACRP, Kenitra	hbenaouda_inra@yahoo.fr
Douaik, Dr Ahmed	INRA, Rabat	ahmed_douaik@yahoo.com
El Hairech, Mr Tarik	DMN, Casablanca	tarik.elhairech@gmail.com
El Hani, Dr Sliman	INRA, Tangier	sliman_elhani@yahoo.fr
Hadria, Dr Rachid	INRA, Oujda	r.hadria@gmail.com
Ismaili, Ms Samira	INRA, Kenitra	ismaili.samira@gmail.com
Tahiri, Mr Mostafa	DSS, Rabat	tahrimostafa@gmail.com
Touil, Mr Mohssine	DSS, Rabat	mohssine71@gmail.com

Annex 2. Exercises

The following Powerpoint presentations were shown:

- MARS Viewer
- The CGMS crop yield forecasting system (also known as "CST introduction and concepts").

Besides, the following exercise material was used which was specially developed in view of the workshop:

- Exercises with the MARS Viewer
- Training course on the CGMS Statistical Toolbox.

The documents on the CGMS Statistical Toolbox / CST have become available also in French in the mean time. The exercise texts are adopted here as part of this report.

A. Exercises with the MARS Viewer

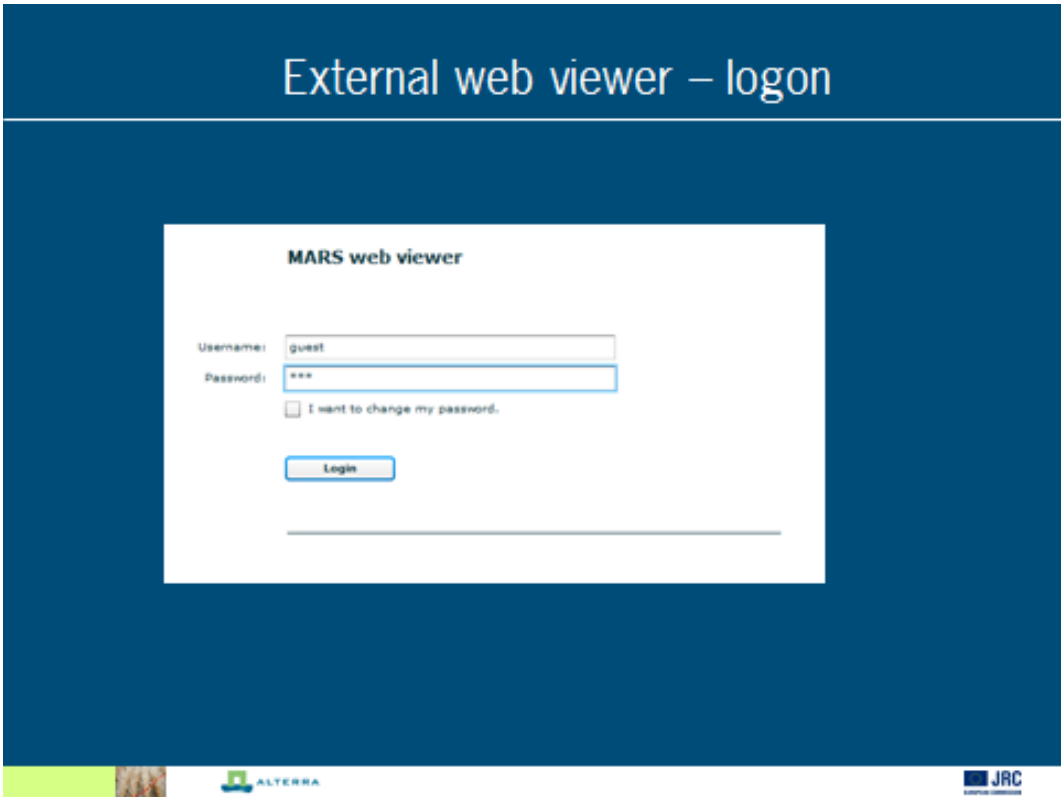


Figure 4: Login page at http://www.marsop.info/glw/bin/webviewer.html

Exercises

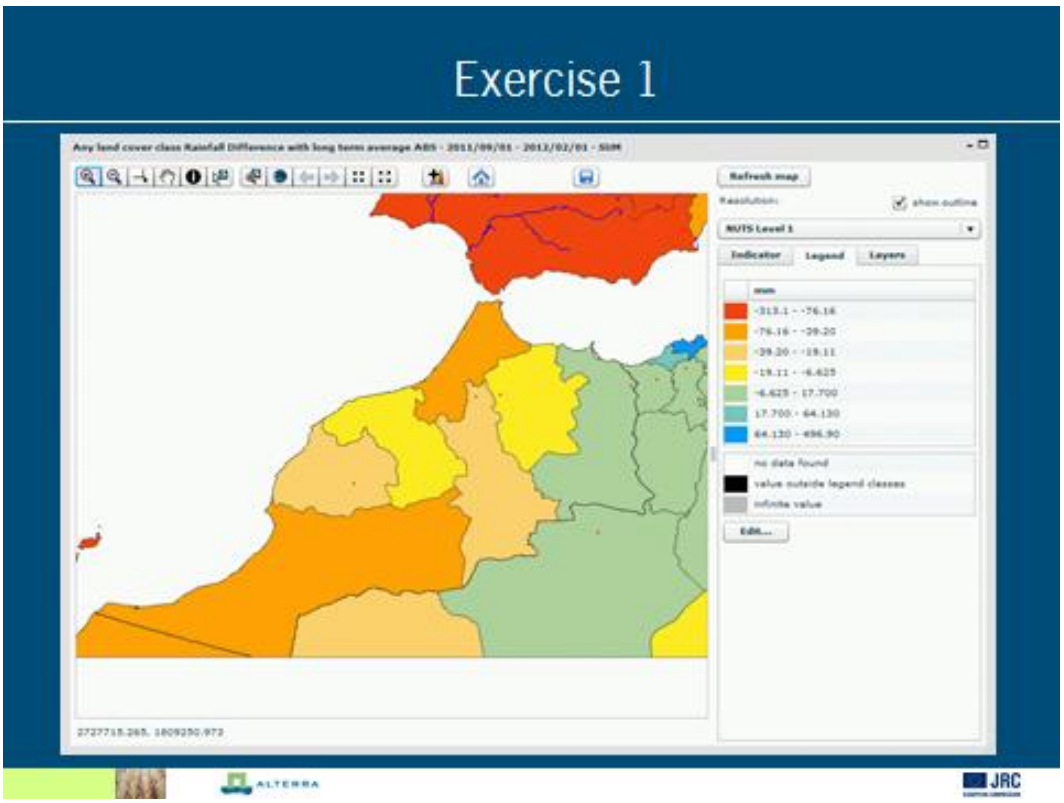
- Select view "map"
- Select region of interest: Europe
- Zoom to: Morocco

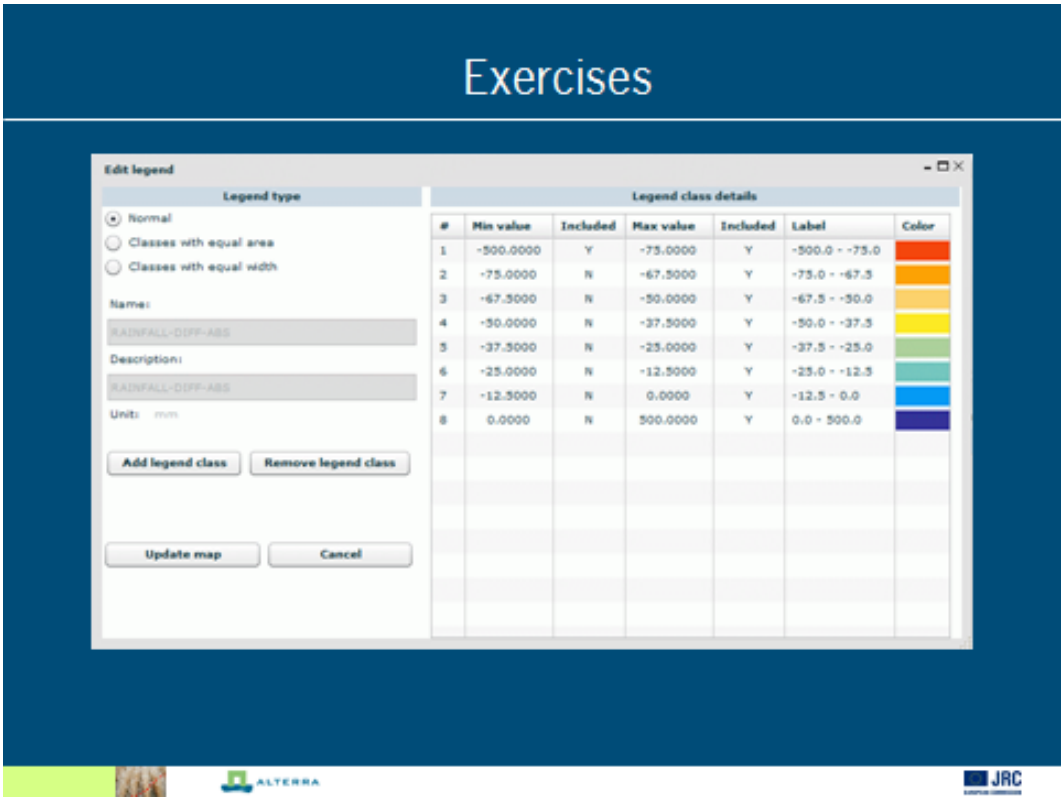
Please wait until all drop down boxes have been filled before continuing

Exercise 1

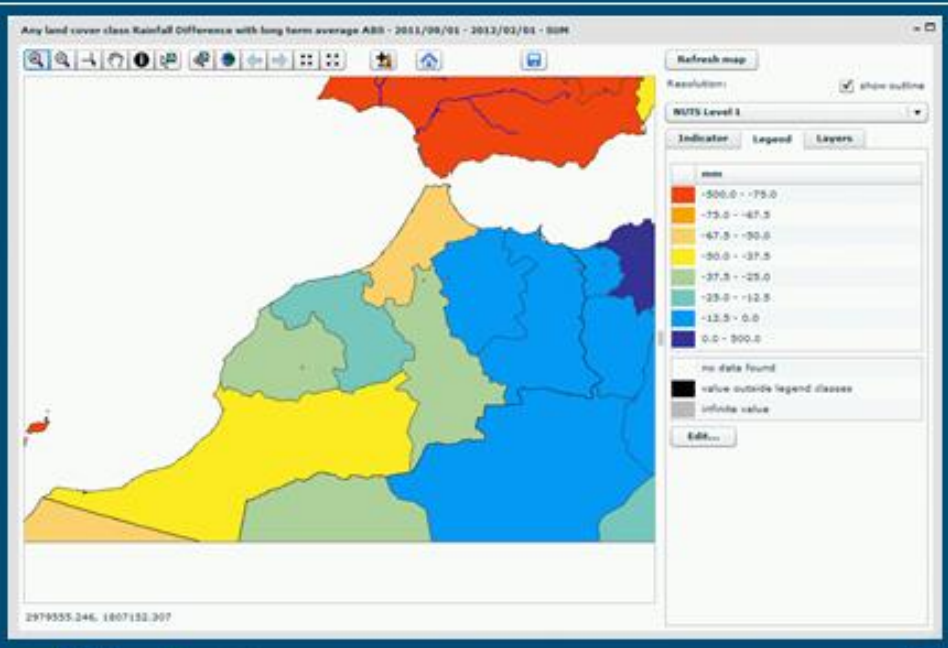
- Open map window
- Select resolution "NUTS level 1" and theme "weather indicators (observed)" ...
- Select the land cover "any land cover class"
- Do the other selections in such way that you will first obtain the rainfall sum over the past period September 1 – February 1, then the absolute difference compared to the long term average
- Refresh the map and look at it
- Change the legend: select the normal legend type. Use the identify tool to establish an appropriate range for Morocco and define 8 new classes, with a class for extreme values at both ends of the range:
  - -500 to -75
  - -75 to ?
  - ...
  - ? to 0
  - 0 to 500

  (NB do not forget to change the labels!!!)
- Update the map


Exercise 1
Any land cover class Rainfall Difference with long term average ABS - 2011/09/01 - 2012/02/01 - SUM
Refresh map
Resolution:
show outline
NUTS Level 1
Indicator
Legend
Layers
mm
-313.1 - -76.16
-76.16 - -39.20
-39.20 - -19.11
-19.11 - -6.625
-6.625 - 17.700
17.700 - 64.130
64.130 - 496.90
no data found
value outside legend classes
infinite value
Edit...
2727715.265, 1809250.972
ALTERRA
JRC


Exercises
Edit legend
Legend type
Normal
Classes with equal area
Classes with equal width
Name:
RAINFALL-DIFF-ABS
Description:
RAINFALL-DIFF-ABS
Unit: mm
Add legend class
Remove legend class
Update map
Cancel
Legend class details
Min value Included Max value Included Label Color
1 -500.0000 Y -75.0000 Y -500.0 - -75.0
2 -75.0000 N -67.5000 Y -75.0 - -67.5
3 -67.5000 N -50.0000 Y -67.5 - -50.0
4 -50.0000 N -37.5000 Y -50.0 - -37.5
5 -37.5000 N -25.0000 Y -37.5 - -25.0
6 -25.0000 N -12.5000 Y -25.0 - -12.5
7 -12.5000 N 0.0000 Y -12.5 - 0.0
8 0.0000 N 500.0000 Y 0.0 - 500.0
ALTERRA
JRC

Exercise 1

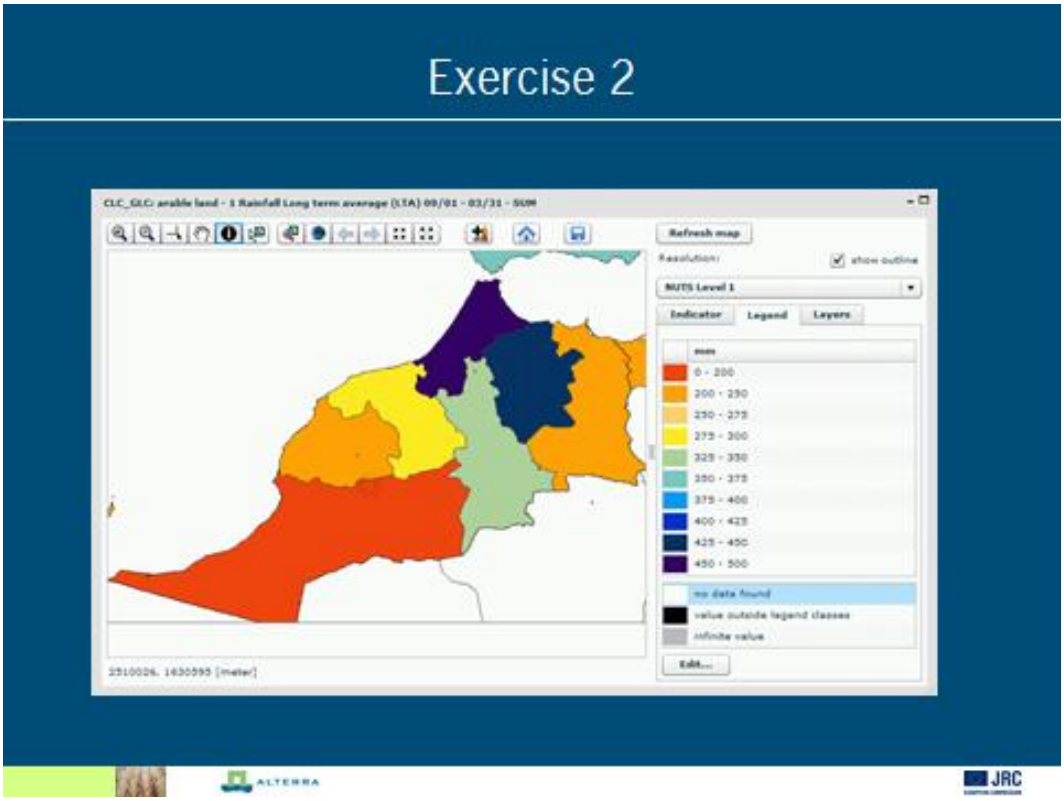


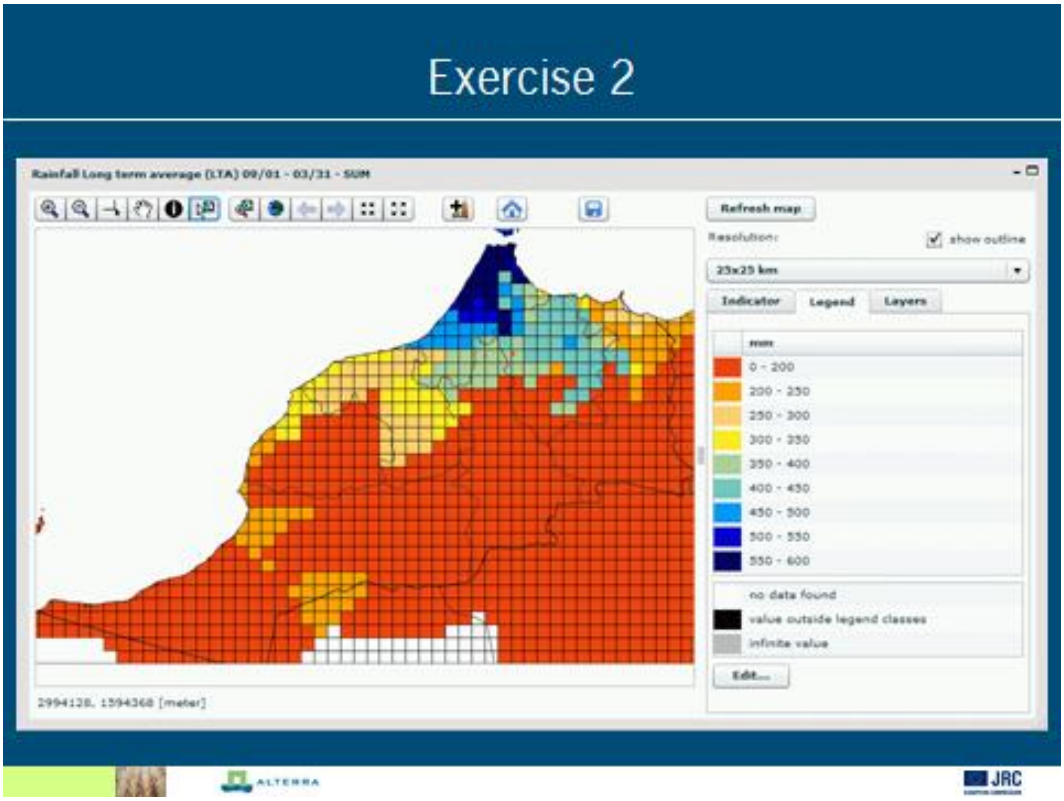
ALTERRA JRC

Exercise 1

- Create another map for the long-term average rainfall sum over the period September 1 thru March 31 – assuming that whatever rain falls afterwards is not relevant for the yield formation anymore. For land cover use CLC-GLC arable land–1.
- Make a similar map at grid resolution
- Compare the two maps and assess how realistic they are.

ALTERRA JRC

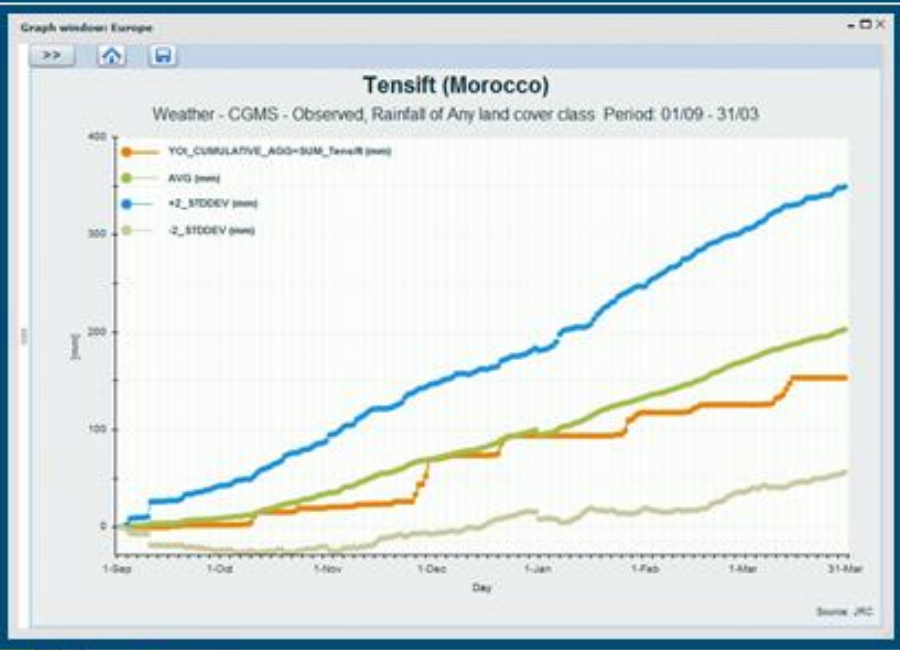

Exercise 2
Refresh map
Resolution:
show outline
NUTS Level 1
Indicator
Legend
Layers
mm
0 - 200
200 - 250
250 - 275
275 - 300
325 - 350
350 - 375
375 - 400
400 - 425
425 - 450
450 - 500
no data found
value outside legend classes
infinite value
Edit...
ALTERRA
JRC


Exercise 2
Rainfall Long term average (LTA) 09/01 - 03/31 - SUM
Refresh map
Resolution:
show outline
25x25 km
Indicator
Legend
Layers
mm
0 - 200
200 - 250
250 - 300
300 - 350
350 - 400
400 - 450
450 - 500
500 - 550
550 - 600
no data found
value outside legend classes
infinite value
Edit...
2994128, 1594368 [meter]
ALTERRA
JRC

Exercise 2

- Go to tab "layers"
- Visualize NUTS level 1
- Visualize the administrative names
- Select region "Tensift" (look for button "activate spatial feature tool")
  Note the selected region will have a red outline!!
- Add a graph window and choose "Analyse one specific year: one indicator ..."
- Select theme "weather...", land cover "any land cover class" and indicator "rainfall" in combination with the cumulative option
- Further select the function type "year of interest"
- Define the period 1 September 2010 – 31 March 2011 (tab "period")
- Refresh graph
- Open the left panel, go to the indicator tab and select the overlapping profile "Confidence interval, average"
- Refresh graph

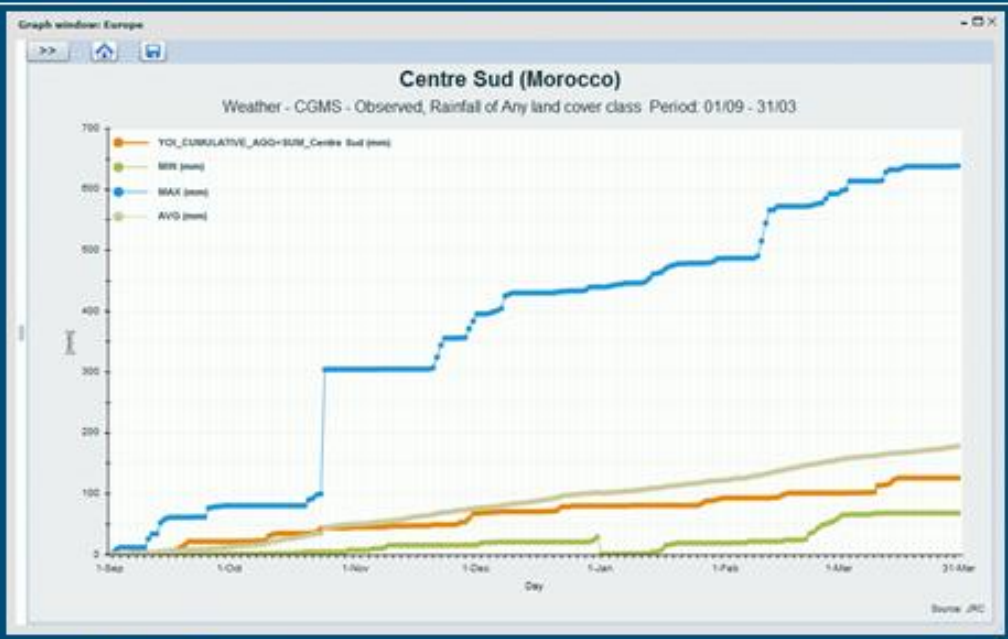
Exercise 2



Exercise 3

- Go to the map window
- Zoom out to the whole region
- Select region "Centre Sud"
- Check the graph window
- Go back to the map window
- Go the indicator tab and select overlapping profiles like closest and most distant years. What are the most closest and most distant year?

Exercise 3



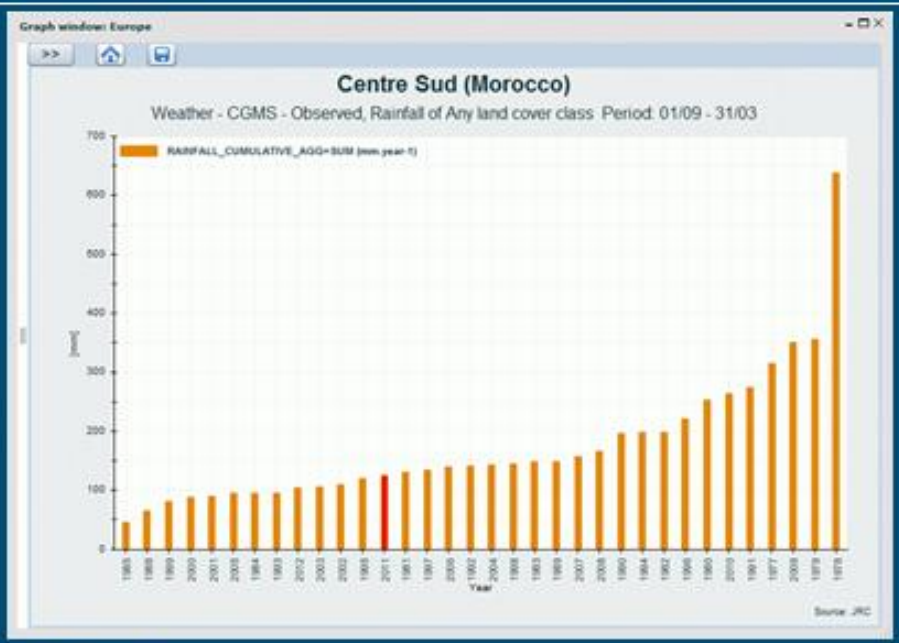
Exercise 3
Centre Sud (Morocco)
Weather - CGMS - Observed, Rainfall of Any land cover class Period: 01/09 - 31/03
Day
ALTERRA
JRC

Exercise 3
Centre Sud (Morocco)
Weather - CGMS - Observed, Rainfall of Any land cover class Period: 01/09 - 31/03
Day
ALTERRA
JRC

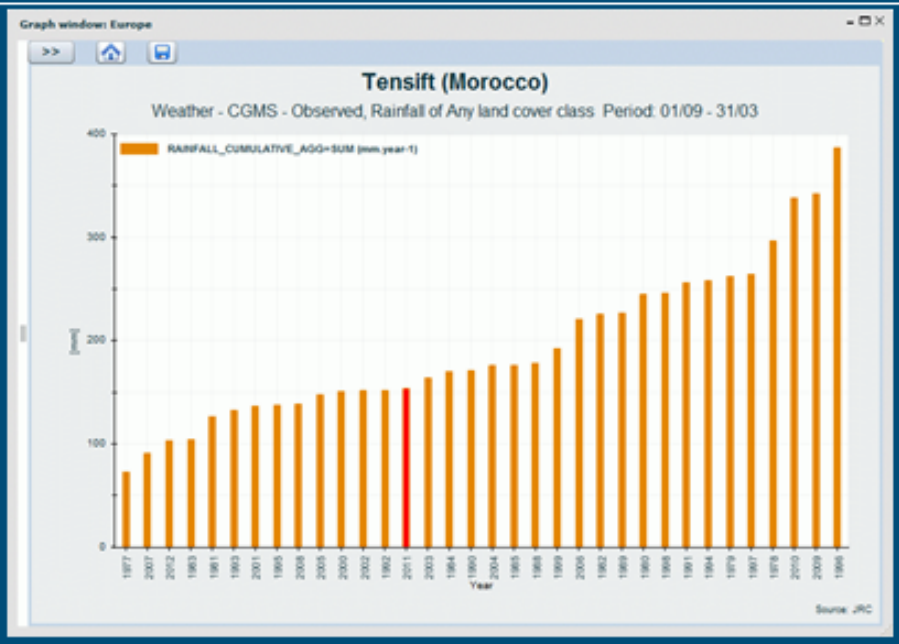
Exercise 4

- Close the graph window and open a new graph window
- Select graph type "analyse all available years"
- Select theme "weather...", land cover "any land cover class" and indicator "rainfall" in combination with the sum aggregation type
- Use the default function type "year of interest"
- Enable the option "order by value"
- Define the period 1 September 2010 – 31 March 2011 (tab "period")
- Refresh graph
- What is the most dry and wet year?
- Do the same for region "Tensift"

Exercise 4

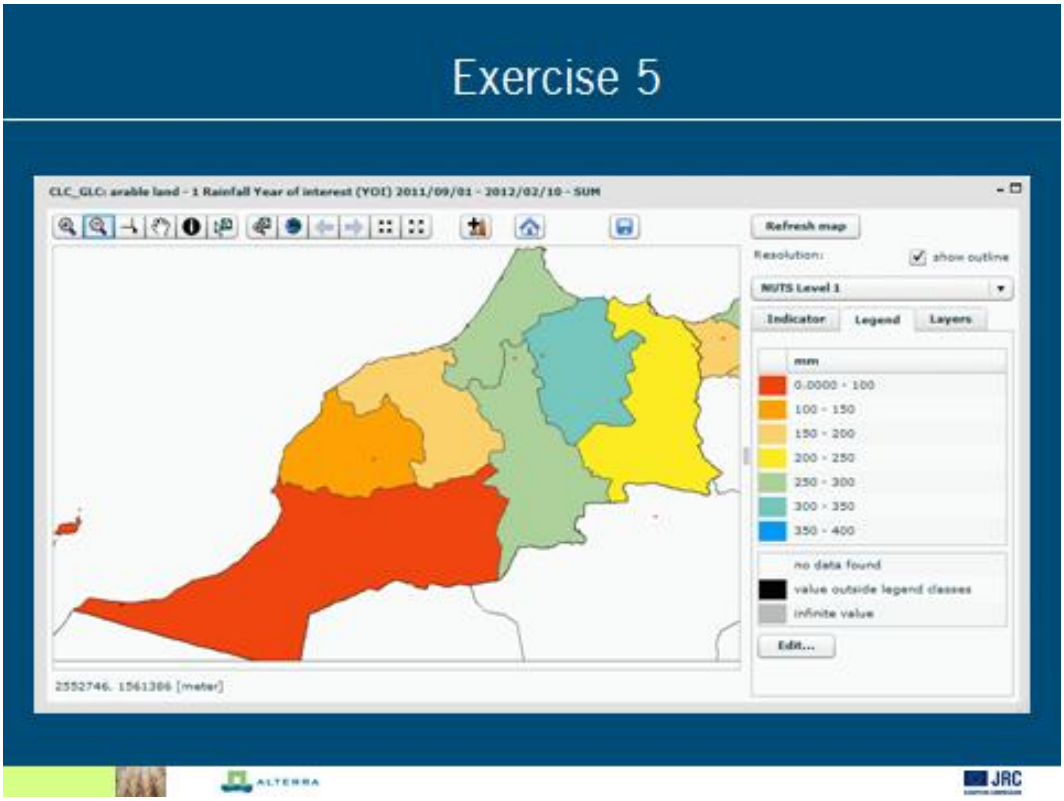


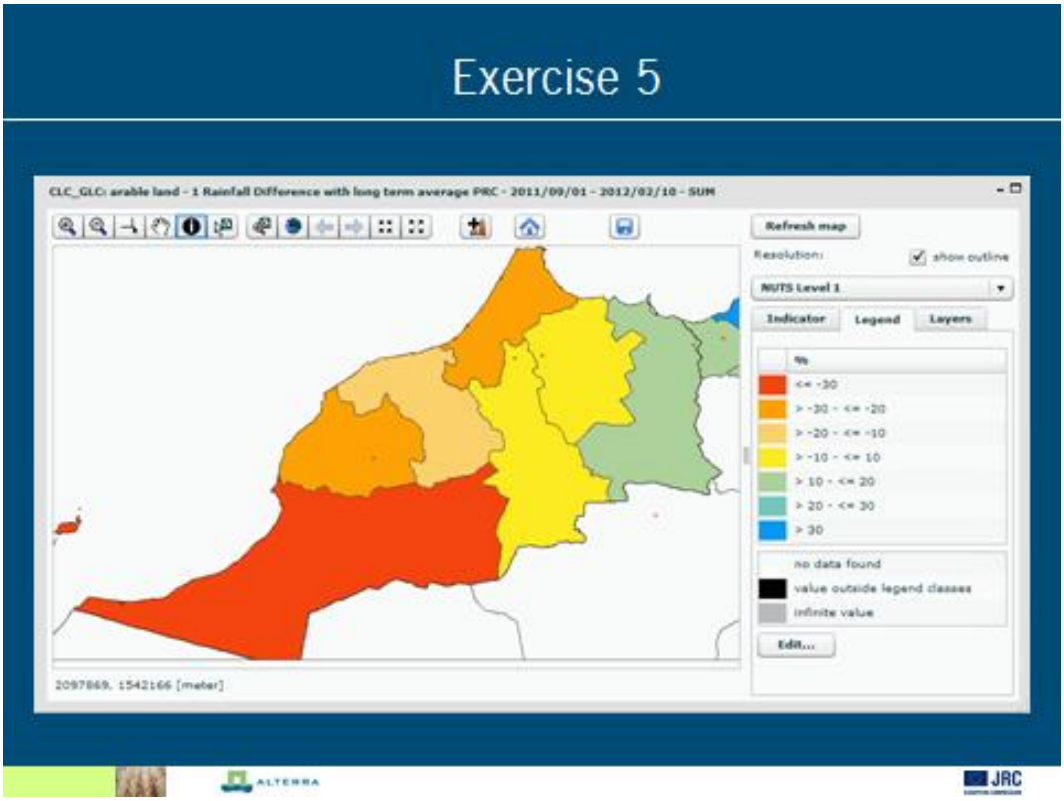
Exercise 4



Exercise 5

- Select NUTS level 1
- Make a map of Morocco showing the cumulative rainfall since September 1 for each region, for land cover CLC_GLC arable land - 1
- Then change the map further to reflect how much the rainfall since September 1 deviates percentually from the long-term average, for each region.

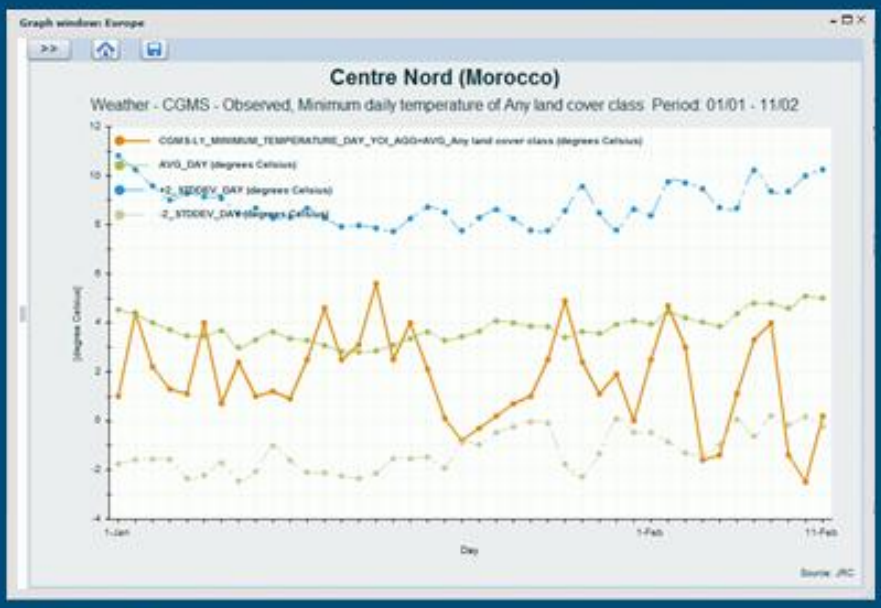

Exercise 5
CLC_GLC: arable land - 1 Rainfall Year of interest (YOI) 2011/09/01 - 2012/02/10 - SUM
Refresh map
Resolution:
show outline
NUTS Level 1
Indicator
Legend
Layers
mm
0.0000 - 100
100 - 150
150 - 200
200 - 250
250 - 300
300 - 350
350 - 400
no data found
value outside legend classes
infinite value
Edit...
2552746, 1561386 [meter]
ALTERRA
JRC


Exercise 5
CLC_GLC: arable land - 1 Rainfall Difference with long term average PRC - 2011/09/01 - 2012/02/10 - SUM
Refresh map
Resolution:
show outline
NUTS Level 1
Indicator
Legend
Layers
%
<= -30
> -30 - <= -20
> -20 - <= -10
> -10 - <= 10
> 10 - <= 20
> 20 - <= 30
> 30
no data found
value outside legend classes
infinite value
Edit...
2097869, 1542166 [meter]
ALTERRA
JRC

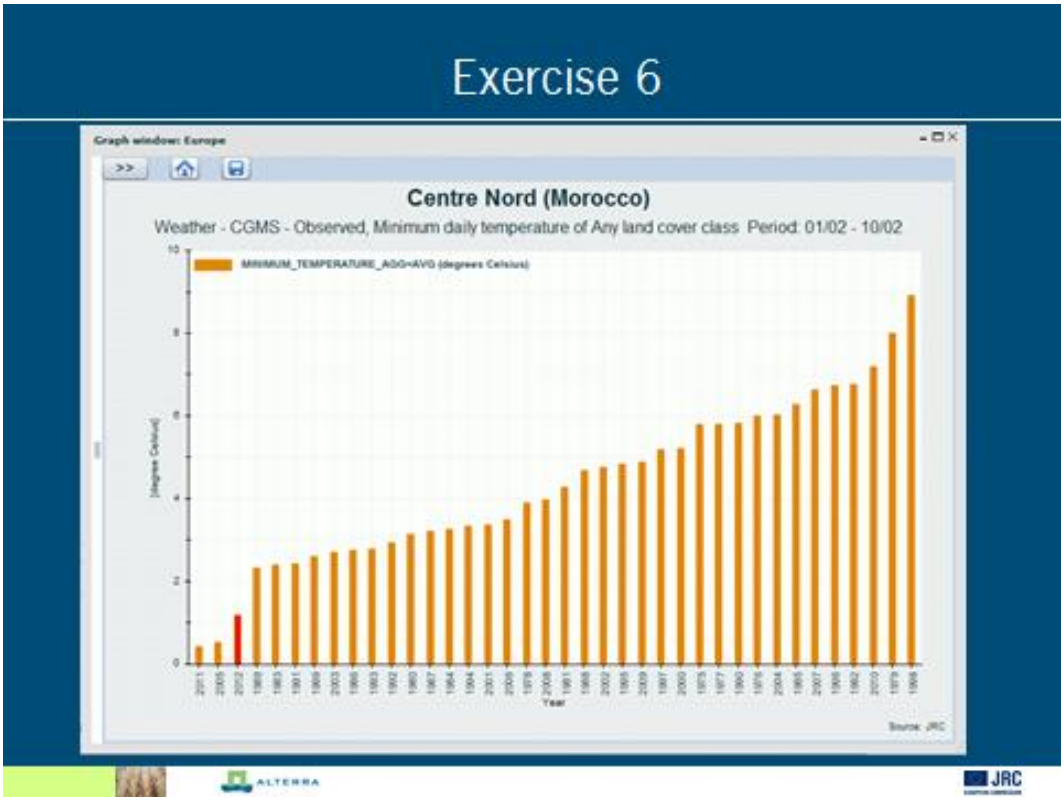
Exercise 6

- Select NUTS level 1 and select the Centre Nord region
- Make a graph for the Centre Nord region of Morocco, showing the minimum temperatures since January 1 with the average and confidence intervals
- Add another line to the graph showing the crop development stage for wheatacc. to the MARSOP3 database
- What do you think about the graph now?
- Can you assess whether the recent cold has affected crop development?
- Add another graph arranging the minimum temperatures for the first decade of February from low to high
- What does last year's bulletin say about low temperatures?

Exercise 6



Exercise 6
Graph window: Europe
Centre Nord (Morocco)
Weather - CGMS - Observed, Minimum daily temperature of Any land cover class Crop - CGMS, Crop development stage
Day
ALTERRA
JRC


Exercise 6
Graph window: Europe
Centre Nord (Morocco)
Weather - CGMS - Observed, Minimum daily temperature of Any land cover class Period: 01/02 - 10/02
MINIMUM_TEMPERATURE_AGG=AVG (degrees Celsius)
Year
ALTERRA
JRC

B. Training course on the CGMS Statistical Toolbox

February 2012
Allard de Wit and Steven Hoek,
Alterra, Wageningen-UR

Introduction

An important component in any regional crop yield forecasting system is the screening and analysis of regional statistical data on crop yield, area and production. Moreover, for yield forecasting purposes it is necessary to understand the inter-annual variability of crop yield and to determine if this variability can be explained by one or more indicators. These indicators can then be used to predict crop yield early in the growing season. Such indicators are often derived from weather data, from satellite information or from crop simulation models.

Within the context of the European MARS system, a dedicated tool was developed that combines these analyses within a single (relatively) easy to use application: the CGMS Statistical Tool (CST). The CST allows 1) to select a region and crop type, 2) select a time period in the year, 3) make a time-series analysis on the available statistical data, 4) carry out a regression or scenario analysis and 5) finally predict crop yield for the target year.

It should be noted that in principle all these analyses can be carried out using modern office tools like Microsoft Excel and there is no “magic” in terms of statistics involved. Nevertheless, carrying out such exercises in Excel is time-consuming, error-prone and cumbersome. Therefore, the CST streamlines the process of analysis and allows to store and document the settings used to make a certain forecast.

During this training course, the different components of the CST will be discussed and illustrated with exercises. Where appropriate, Microsoft Excel will be used to visualize and demonstrate the logic behind some of the analysis in the CST.

Note: This document does not describe installation procedures related to the CST. For such information see: http://e-agri.wikispaces.com/CGMSStatTool

Part 1: Time trend analysis

Time trend analysis is used to analyse the reported yields for a certain region and make an assessment of:

- data availability
- existence of anomalous years that might be excluded from the analysis (for example an error in the statistical data)
- existence of trends in the time-series

The time trend analysis window has a lot of functionality that can be used to interpret the regional yield statistics (figure 1). The different elements in this window will be discussed during the exercises with the CST.

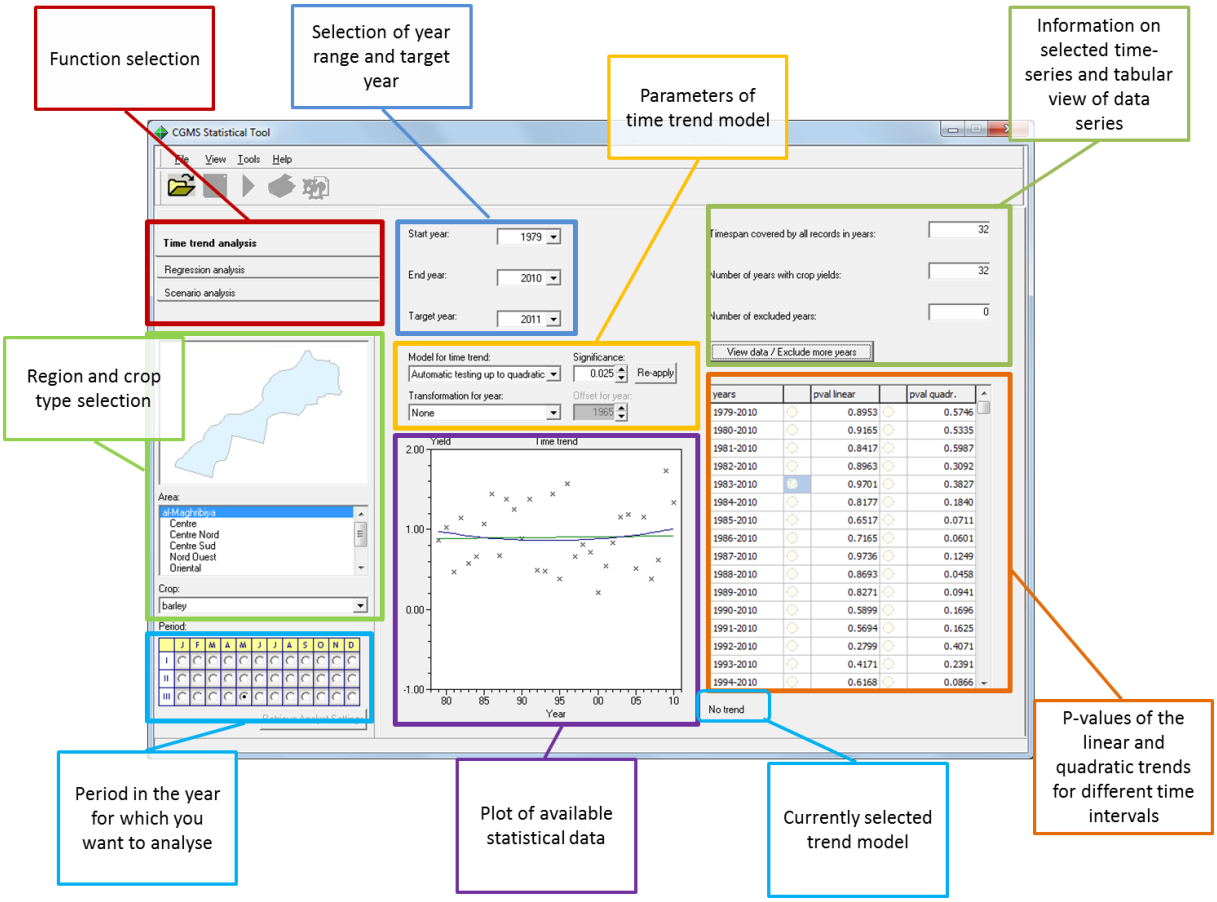


Figure 1. Explanation of functionality in the time trend analysis window.

Data availability

First of all we will start with a simple exercise to determine what statistical data is available in the database. For this purpose we will not select some regions and crops and let CST report the information.

Exercise 1.1: analysing data availability

Note in the table below the available start year, end year and the number of missing years for the region "Al-maghribiya" (whole of Morocco). Next, select the region "Centre" and carry out the same analysis.

	Al-Maghribiya			Centre		
	start	end	missing	start	end	Missing
Barley						
Durum wheat						
Soft wheat						

Anomalous points detection

The chart window in the centre of the CST shows the time-series of regional yield statistics that are available for the selected crop type and region. This chart of yield statistics versus time is particularly useful for identifying trends and anomalous values in the yield statistics. Anomalous values can be regarded as points that are outside the normal distribution of yield values that can be expected for the region. CST does not provide statistical tools for outlier detection, but relies on visual interpretation.

Exercise 1.2: Searching for anomalous points in the statistics

Note in the table below the region, year of detected anomalies for soft-wheat in the regions Centre, Oriental and Tensift

region	year	Reason for outlier

Extended data availability (Data view)

Having found some points that are looking anomalous, we would like to exclude the points from the analysis. For this purpose we can use the option "View data/Exclude more years" (see figure below). The view data window will give you a complete overview of the available statistical data and the available indicators for the selected region, crop and period. Particular years can be manually excluded using the selection buttons on the left side.

Exercise 1.3: Exclude data using the "View data" window

Select the region "Tensift", select the crop "soft-wheat" and Period "March-III". Next identify the years with the anomalous statistical data, then open the "View data/exclude more years" window (see figure 2). Than deselect the anomalous years and click OK.

Q2: What happens to the chart and the selected trend model?

Data view

Below available data are listed for the selected area, crop and decade. You can exclude data for particular years by unchecking the boxes on the left.

Nuts area: Tensift — Decade: 9

Crop: soft wheat — Included years: 1979-2010

Included	Year	Official Yield	01 Potential Above Ground	02 Potential Storage Organs	03 Water Limited Above Ground	04 Water Limited Storage Organs	05 Cum. rainfall since Septembe	06 Cum. NDVI since February	07 Cum. DMP since March 1
☑	1979	0.721594	11787.964	5939.062	10689.943	4982.572	272.300000	*	*
☑	1980	0.525941	11988.237	5526.602	11723.108	5371.871	247.100001	*	*
☑	1981	0.269628	10813.477	3875.804	8558.927	2372.02	128.799999	*	*
☑	1982	-2.15	11813.458	4652.595	10831.373	3978.712	226.600001	*	*
☑	1983	0.081163	11117.575	4654.936	9848.409	3598.749	113.799999	*	*
☑	1984	0.332661	10749.785	2810.6	10425.934	2742.435	176.599998	*	*
☑	1985	0.85443	*	*	*	*	181.300003	*	*
☑	1986	0.669692	*	*	*	*	*	*	*
☑	1987	0.301221	*	*	*	*	*	*	*
☑	1988	1.133548	12026.058	5358.03	11344.476	4780.193	180.999999	*	*
☑	1989	0.976241	12011.984	5374.384	10553.23	4128.772	242.700001	*	*
☑	1990	0.379343	11990.776	5681.756	9001.482	3443.149	187.400000	*	*
☑	1991	1.215346	12317.025	5221.022	12246.992	5178.406	272.400001	*	*
☑	1992	0.305233	12475.766	5859.186	9041.195	3246.892	151.399999	*	*
☑	1993	0.337528	11885.02	4669.899	10341.844	3333.468	145.099999	*	*
☑	1994	1.275553	12251.932	5278.431	11618.299	4875.791	273.499997	*	*
☑	1995	6.3	11716.707	5515.98	10382.909	4554.819	140.400000	*	*
☑	1996	1.510029	10455.768	4743.726	10455.511	4743.555	428.599999	*	*
☑	1997	0.562937	11229.337	5808.212	8529.454	3501.342	281.399999	*	*
☑	1998	0.822337	10725.149	5582.79	8707.756	3856.565	251.000000	*	*
☑	1999	0.58903	11936.194	4474.229	11637.93	4372.172	195.800000	2.25699996	164.800003
☑	2000	0.479472	11420.914	5583.448	7125.838	1976.869	160.900000	2.03699998	94.8999977
☑	2001	0.391453	11751.829	5445.003	7766.639	2134.99	151.999999	1.98000001	115
☑	2002	0.714092	11138.791	5800.017	7995.715	3280.911	160.499998	1.84500002	110.599996
☑	2003	1.022235	11434.853	4998.476	8621.396	2885.072	183.100001	2.71799999	178.400001
☑	2004	1.019913	12145.406	6114.538	9268.171	3738.462	193.600000	2.71700003	196.799995
☑	2005	0.458007	10485.973	3811.855	8516.421	2440.144	149.399999	1.92199999	115.199998
☑	2006	1.112434	11126.844	4067.052	10129.036	3191.174	234.200001	3.13900005	270.300003
☑	2007	0.390482	11865.648	6095.243	8240.532	2874.027	97.6000008	1.70700001	108.200000
☑	2008	0.383945	11653.842	6394.938	7965.193	3226.775	142.000001	2.20800000	130.799999
☑	2009	2.350245	11353.224	4815.178	10671.282	4326.489	358.100002	3.24900007	246.5
☑	2010	1.422196	11180.329	5736.517	10950.964	5507.363	354.200000	2.87700003	227.400001
Target	2011	*	12006.734	5535.372	9420.148	3390.789	190.000001	2.54899999	175.200004

Exclude missing | Reset | Copy to clipboard | Cancel | OK

Figure 2. the data view window for the region Tensift.

Trend selection

Upward trends are often occurring in the time-series of regional crop yields, which are often the result of improvements in technology (better crop varieties, better fertilizers, better crop management. Downward trends are more rare, but are also occurring, for example, due to soil degradation.

Such trends need to be taken into account when forecasting crop yield because they usually cannot be explained by the crop yield indicators. For example, in the Netherlands, yield of wheat has increased from 5.72 ton/ha over the period 1976-1980 to 8.54 ton/ha over the period 2000-2005. This increase cannot be explained by changes in weather and is solely the result of the "technology trend".

The CST provides some advanced tools for trend model selection:

- A linear and quadratic trend is automatically fitted in the chart window over the selected time-period.
- With the option "automatic testing up to quadratic" the best trend model (given the significance value) is automatically selected and displayed. See figure 1, the field "Currently selected trend model".
- In the lower right panel, all alternative time-periods for trend selection are listed with the p-values of the linear and quadratic trend over that period.

Together these tools can be used to determine if there is a trend and if so, what the best period is for trend selection. Finally, a trend can be forced into the system. This is useful if you suspect that there is a trend but it is masked by a single point at the end/beginning of the time-series.

Exercise 1.4: Is there a trend?

Select the region "Al-Maghribiya" and subsequently select the three crop types.

Q1: Note down the default trend selected (e.g. None, Linear, Quadratic) and look through the list of other periods for trend selection. Select the period with the most significant trend (lowest p-value)

Crop	Default model	Period with most significant trend and p-value.
Barley		
Durum wheat		
Soft wheat		

Q2: Is there a trend for the three cereal crops for Morocco (as a whole):

Part 2: Regression analysis

Regression analysis is one approach in finding a relationship between the inter-annual variability in the crop yield statistics on the one side, and the indicators (weather, crop simulation, satellite) on the other hand. When such a relationship is found and parameterized it can be used to forecast the crop yield for the target year.

Indicators tab

After selecting the "Regression analysis" tab, the first stage is to select indicators from the "Available indicators" selection box

Exercise 2.1: Indicator Selection

Select region "Centre", crop soft-wheat and period J-III (third decade of January). Leave the Start/End year on 1979-2010. Now select the regression tab.

Q1: How many indicators are available in the "Available indicators" selection box?

Now select the period first decade of February (F-I).

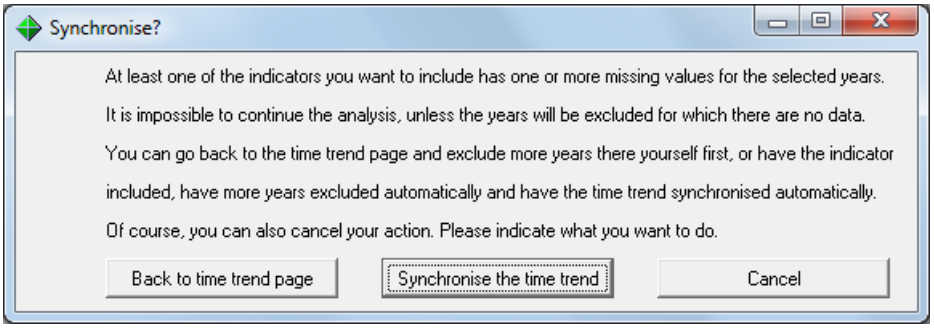
Q2: How many indicators are now available in the "Available indicators" selection box. Which indicator is now available and why?

Now select the period: first decade of March (M-I).

Q3: How many indicators are now available in the "Available indicators" selection box? Which indicator is now available and why?

Q4: Some indicators are listed as having missing values. Particularly the satellite indicators have many missing values. Can you explain this? Tip: have a look at the "Data view" window in the trend analysis.

Add all indicators in the "available indicators selection box" into the box with "Free indicators" using the **>>** button. (See page 25 for an explanation of Free or Forced indicators). When adding all indicators, the CST will come up with the following window:



This is needed because all indicators and official yield need to be present in the analysis. Press the "Synchronize" button to have all indicators synchronized over the same years. The CST can also compute the correlation matrix for you between the different indicators. An indicator that is highly correlated (large absolute coefficient of correlation) with another indicator has little additional value in the regression analysis. A quick look on the correlation matrix will show you that all CGMS indicators (01..04) are highly correlated. The sum of rainfall (05) and satellite derived indicators (06, 07) are much less correlated.

Options tab

The options tab lists a fairly large number of options for regression models (figure 3). The most fundamental choice to make here is to use either "single free indicator" or "Best

subset selection". In the former case only one indicator at a time is used to build the regression model, in the latter case the CST will test many combinations of indicators to find the best fitting model.

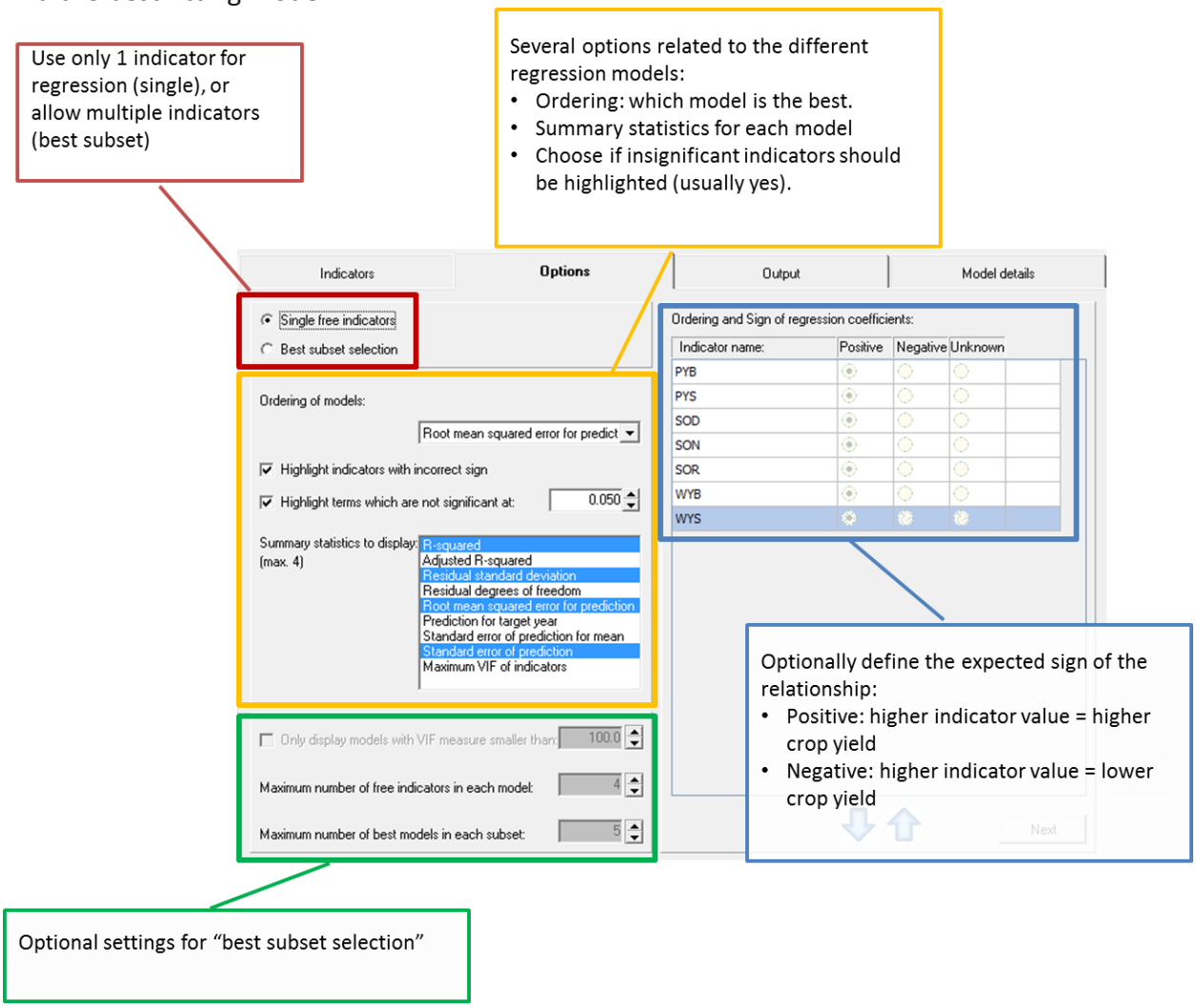


Figure 3. Explanation of functionality in the regression options window.

Exercise 2.2: Single Free indicator analysis
Settings in the time trend analysis window:

- Region: Centre
- Crop: soft-wheat
- period: March-I (first decade of March).
- Start/End year: 1999-2010 (to have no missing data).

Continue in the Regression Tab:

- Add all indicator to the Free Indicators

Continue in the Options Tab:

- Select "Single Free indicator"
- Mark all indicators as having a positive sign
- Leave the rest default

Continue to the Output Tab
The table shows you the results of the regression models for the different indicators.
Q1: How many regression models are listed and what is the *None* model?
Q2: What is the best model and what is the R^2 values of this model?
Q3: A measure of the error on the crop yield prediction can be obtained from the Root Mean Squared Error (RMSE) for prediction. What is the RMSE for the best model and how does it compare to the *None* model.
Q4: Some of the cells with the T-values are marked either orange or red, what does this mean?
By clicking on the number of each model a detailed report of the model is provided.
Q5: What is the predicted crop yield (Prediction for target year) and it standard error (Standard Error of Prediction (New))?

Exercise 2.3: Single Free indicator analysis
Repeat exercise 2.2, but set the period to April-III
Q1: What is the best model and what is the R^2 values of this model?
Q2: What is the RMSE for the best model and how does it compare to the *None* model.
Q3: What is the predicted crop yield (Prediction for target year) and it standard error (Standard Error of Prediction (New))?
Q4: Is the predicted yield very different from the predicted yield in March-I

We will now repeat exercise 2.2/2.3 with the "Best subset selection". Using the "Best subset selection" we can try to build a better regression model and obtain a better forecast with smaller error bounds.

Exercise 2.4: Best subset selection
Repeat exercise 2.3, (set the period to April-III). Continue in the Options Tab:

- Select "Best subset selection"
- Mark all indicators as having a positive sign
- Leave the remaining settings default

Continue to the Output Tab
Q1: What is difference in the ordering of models and why?
Q2: What is the best model with a single indicator? How does this compare to the result you found with Question 1 of Exercise 2.3?
Q3: Is there a better model with two indicators, with three, with four?

We will now put the knowledge gain in the previous exercises to practice in the sense that we will be making a "forecast" (actually a "hindcast" as it is already in the past) for the 2011 season using the data in the CST.

Exercise 2.5: Make a forecast for 2011 for soft-wheat in region Centre
For this exercise you will be using the excel spreadsheet "Forecast_centre2011_training.xlsx".
Settings in the time trend analysis window:

- Region: Centre
- Crop: soft-wheat
- period: January-I
- Start/End year: 1999-2010 (to have no missing data).

Continue in the Regression Tab:

- Add all indicator to the Free Indicators

Continue in the Options Tab:

- Select "Best subset selection"
- Mark all indicators as having a positive sign
- Leave the rest default

Exercise 2.5: continued

Continue to the Output Tab

- Select the best model and click on the model to get the model details and enter the following details in the Excel spreadsheet:
  - The model name (e.g. 05, or 05+06, etc)
  - The Prediction for target year under the column "forecasted yield"
  - Standard Error of Prediction (New) under Standard Error

Repeat the analysis for all decades until decade 15 (May-III) and have a look at the resulting chart:
Q1: What happens to the uncertainty on the forecast during the growing season?
Q2: What seems to be the best decade in the season to make the forecast?

Part 3 – Scenario analysis

Regression analysis is a well-known tool for predictive model building. Nevertheless, regression analysis relies on linearity of the relationships between the depending and the independent variables which may not be true. Scenario analysis is an alternative for regression analysis and can be particularly useful when regression analysis fails to provide any model with reasonable explainory power or for forecast years with extreme events. Note however that the scenario analysis provides less information about standard error and uncertainty than regression analysis.
The principle behind scenario analysis is to look for years in the past that are similar to the current growing season and is based on a principle component analysis. This approach makes less assumptions on the distribution of the data and simply computes a similarity score. Given a user-defined threshold, the *n* most similar years are than selected and the forecast for those years is used to make a forecast for the current year.
Figures 4 and 5 provide a short explanation of the functionality under the Options and Output tab.

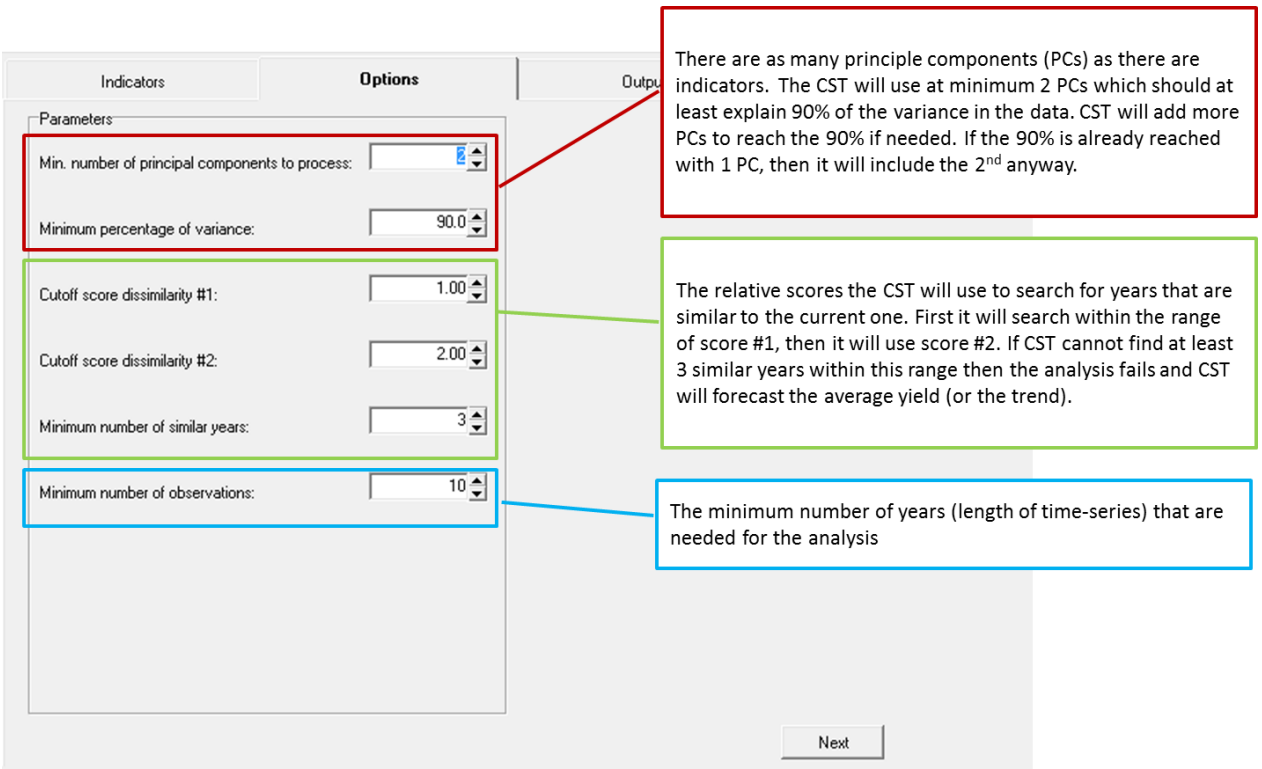


Figure 4. Explanation of functionality under the Options Tab

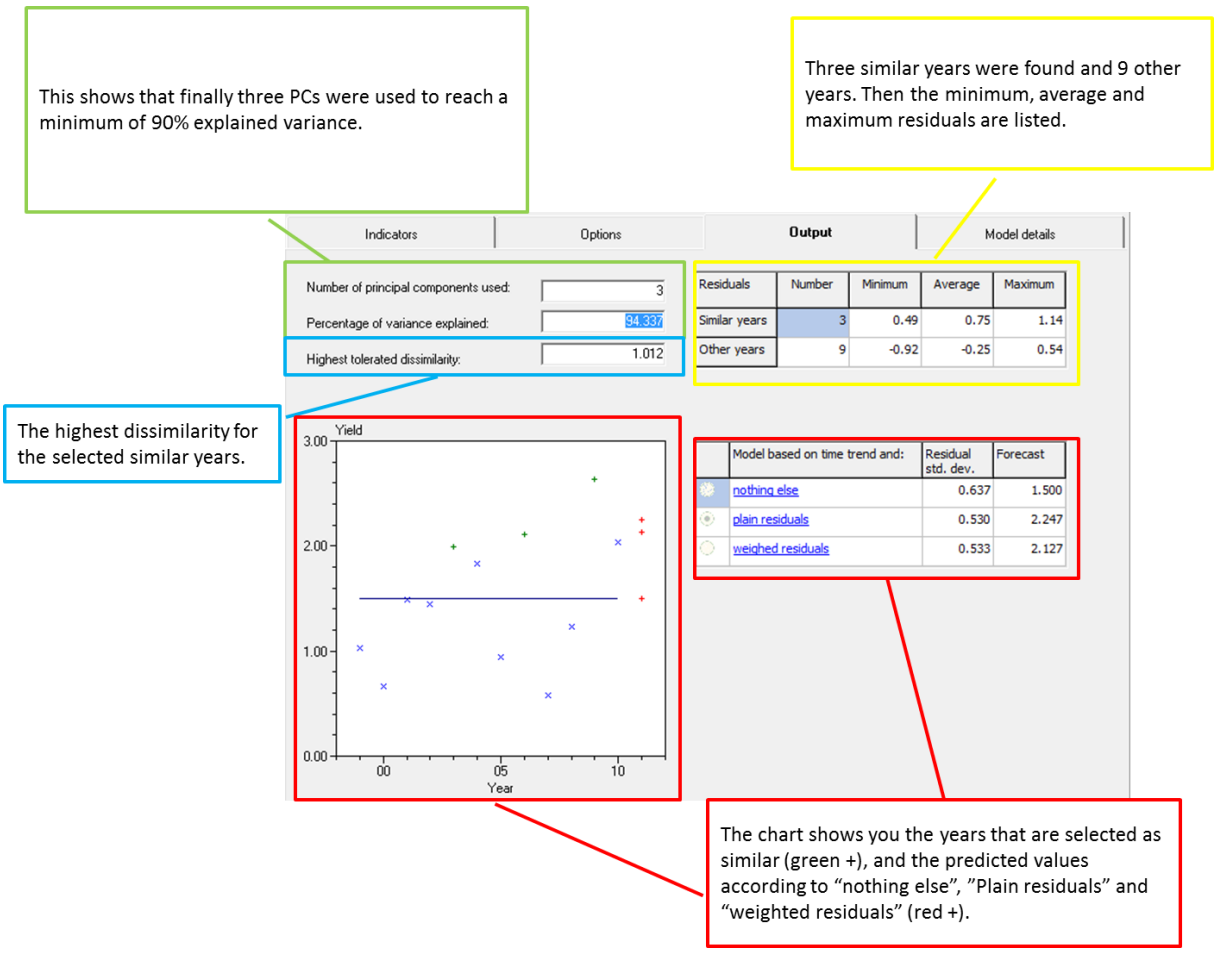


Figure 5. Explanation of results under the Output Tab.

Exercise 3.1: Scenario analysis

Settings in the time trend analysis window:

- Region: Centre
- Crop: soft-wheat
- period: May-III
- Start/End year: 1999-2010 (to have no missing data).

Continue in the Scenario Analysis Tab:

- Add all indicator to the Free Indicators

Continue in the Options Tab:

- Set the minimum number of similar years to 3
- Leave remaining options default

Continue to the Output Tab

The table shows you the results of the Scenario analysis.

Q1: Which years are selected as similar years, does it make sense?

By clicking on the number of each model a detailed report of the model is provided.

Q2: What is the predicted crop yield (Prediction for target year) and its standard deviation

Exercise 3.1: Scenario analysis
Settings in the time trend analysis window:

- Region: Centre
- Crop: soft-wheat
- period: May-III
- Start/End year: 1999-2010 (to have no missing data).

Continue in the Scenario Analysis Tab:

- Add all indicator to the Free Indicators

Continue in the Options Tab:

- Set the minimum number of similar years to 3
- Leave remaining options default

Continue to the Output Tab
The table shows you the results of the Scenario analysis.
Q1: Which years are selected as similar years, does it make sense?
By clicking on the number of each model a detailed report of the model is provided.
Q2: What is the predicted crop yield (Prediction for target year) and its standard deviation

Exercise 3.2: Yield forecast with scenario analysis
For this exercise you will be using the excel spreadsheet "Scenario_Forecast_centre2011_training.xlsx".
Settings in the time trend analysis window:

- Region: Centre
- Crop: soft-wheat
- period: January-I
- Start/End year: 1999-2010 (to have no missing data).

Continue in the Scenario Analysis Tab:

- Add all indicator to the Free Indicators

Continue in the Options Tab:

- Set the minimum number of similar years to 3
- Leave remaining options default

Exercise 3.2: continued

Continue to the Output Tab

- If available select the model with "weighted residuals", otherwise use "nothing else". Next click on the model to get the model details and enter the following details in the Excel spreadsheet:
    - The Prediction for target year under the column "forecasted yield"
    - Pessimistic value
    - Optimistic value

Repeat the analysis for all decades until decade 15 (May-III) and have a look at the resulting chart:

Q1: What happens to the uncertainty on the forecast during the growing season?

Q2: If you compare the results from the scenario analysis to the results from the regression analysis, which method do you prefer and why?

Answers to the questions

Exercise *1.1*

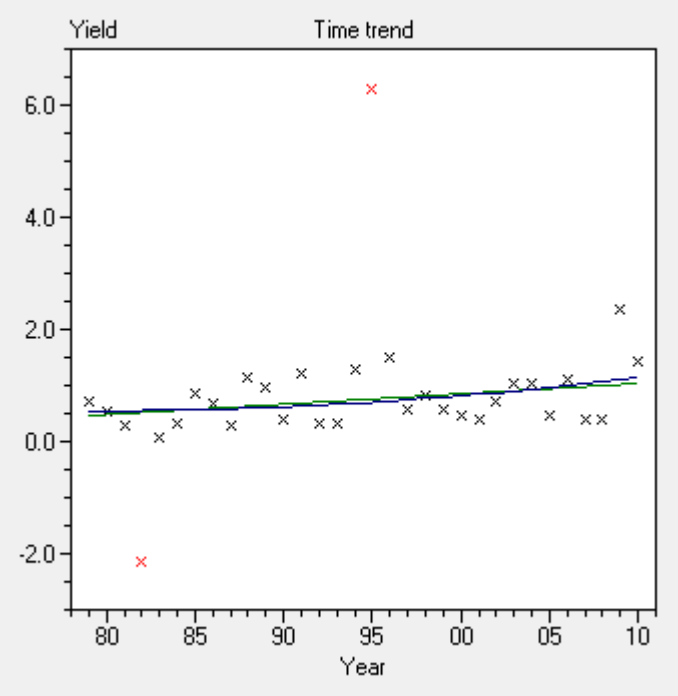
	Al-Maghribiya			Centre		
	start	end	missing	start	end	Missing
Barley	1979	2010	5	1980	2010	3
Durum wheat	1979	2010	1	1988	2010	1
Soft wheat	1979	2010	3	1979	2010	0

Exercise *1.2 (crop=Soft-wheat)*

region	year	Reason for outlier
Tensift	1982	Yield value is negative
Tensift	1995	Yield value is more than 5 times greater than the average yield.

Exercise 1.3: Exclude data using the "View data" window

The deselected years with anomalous points (1982 and 1995) are marked with red crosses in the chart (see below). There is no impact on the selected trend: still there is no trend.



Exercise 1.4: Is there a trend?

Crop	Default model	Period with most significant trend and p-value.
Barley	No trend	1997-2010, p=0.0894 (not significant at 0.025)
Durum wheat	No Trend	1999-2010, p=0.0467 (not significant at 0.025)
Soft wheat	No trend	1999-2010, p=0.0975 (not significant at 0.025)

Is there a trend for the three cereal crops in Morocco: Clearly no!

Exercise 2.1: Indicator Selection

Q1: 5 indicators, four coming from CGMS (01..04) and one derived directly from the meteorology (05 accumulated rainfall)

Q2: One additional indicator is available "06 Sum of NDVI", this indicator is calculated from the 1st of February and thus is not available in January.

Q3: One additional indicator is available "07 Sum of DMP", this indicator is calculated from the 1st of March and thus is not available in February.

Q4: The data view window will tell you that the satellite indicator only start to become available in 1999. This is because the SPOT-VGT satellite was launched in 1998. For the CGMS indicators some years are missing for unknown reason.

Exercise 2.2: Single Free indicator analysis, March-I

Q1: 8 models are listed: you have 7 indicators plus the so-called *None* model which is the make using only the average or the trend for yield prediction.

Q2: The best model is model 05 (sum of rainfall) with an R^2 of 76.59%

Q3: the RMSE of model 05 is 0.40 ton/ha, which is considerable better than the None model which has an RMSE of 0.66 ton/ha.

Q4: The models whose T-value is marked orange are not significant at the given significance level. "Red" models are both not significant and have an wrong sign (negative T-value)

Q5: Predicted yield is 1.809 ton/ha with a standard error of 0.341 ton/ha

Exercise 2.3: Single Free Indicator analysis, April-III

Q1: Best model is model 06 (sum of NDVI) with an R^2 of 80.24%, but the model 05 is close.

Q2: RMSE is 0.32 ton/ha, which is much better than the None model (0.66 ton/ha)

Q3: The predicted crop yield by "model 06" is 2.218 ton/ha and its standard error is 0.329 ton/ha

Q4: Predicted yield is quite a bit higher than in March and the standard error is a bit lower. Probably weather conditions where favourable and the crop developed better than average.

Exercise 2.4: Best subset selection

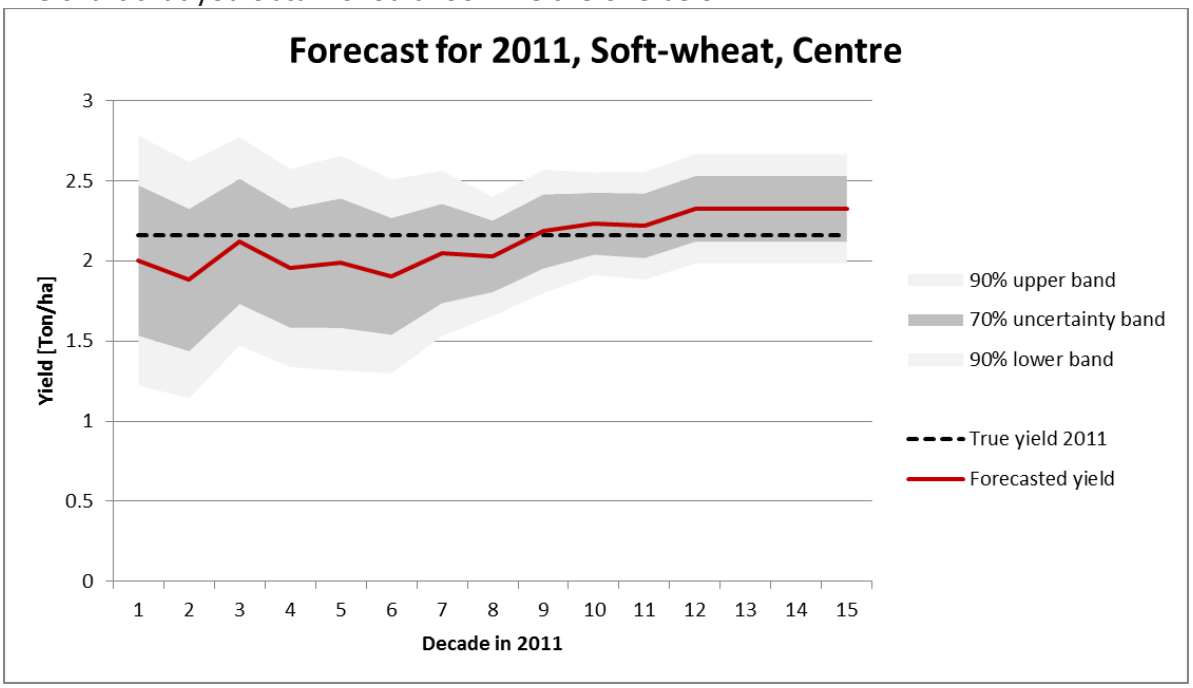
Q1: Models are now ranked according to the number of indicators they have, the second ordering it according to model performance

Q2: The best model with 1 indicator, is the same as the best model using the Single Free Indicator approach (Indicator 06)

Q3: There are two models with two indicators that are valid: “05-07” en “05-06”. the first one (“05-07”)has slightly lower error than the second. There no other valid models, because all other models contain indicators that do not significantly contribute to the regression model. In other words, they all have one or more indicators that are not significant (orange or red cells).

Exercise 2.5 making a forecast

The chart that you obtain should look like the one below:



Q1: clearly the uncertainty is decreasing as the growing season progresses because the models are capable of explaining more variability in the historic reported crop yields (R^2 is increasing with decades).

Q2: The best point in the growing season is actually the point where the standard error is the lowest. In this case this would be decade 10, after decade 10 the standard error is again increasing slightly.

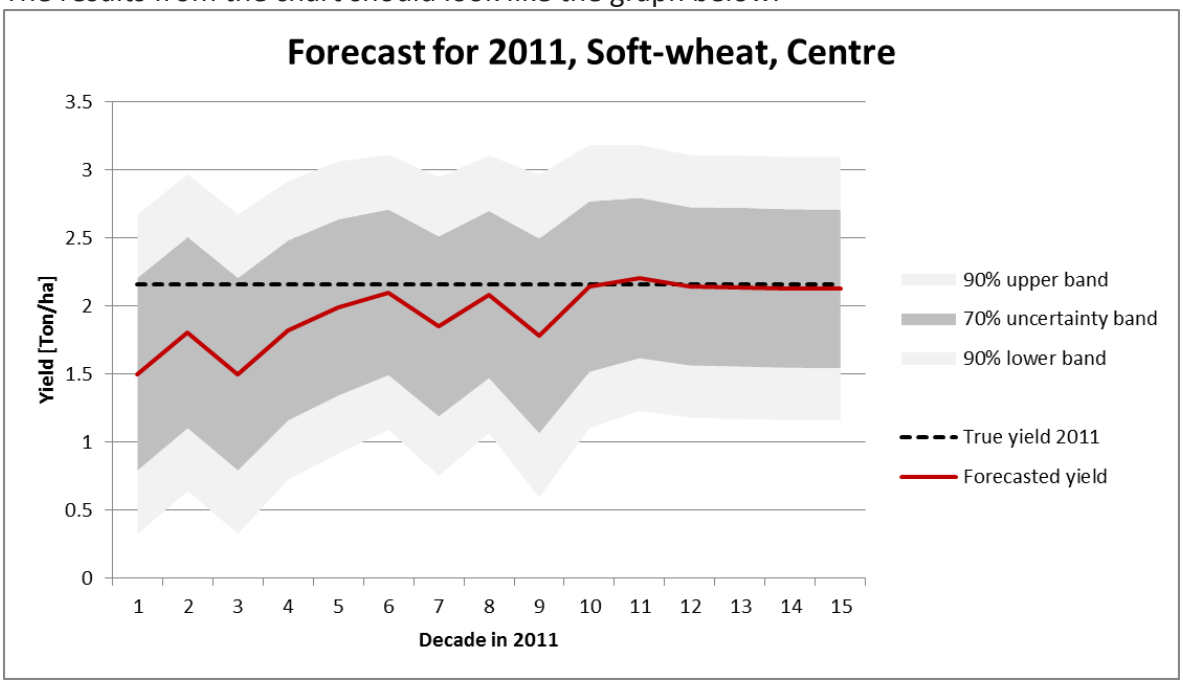
Exercise 3.1

Q1: The years 2003, 2006 and 2009 are selected as similar years. These are all years where the yield is well above average, so yes it makes sense.

Q2: The forecasted yield is 2.127 ton/ha with a standard deviation of 0.533 ton/ha.

Exercise 3.2

The results from the chart should look like the graph below.



Q1: It is quite striking that, although the forecast converges to the true value, the uncertainty bounds on the forecast do not narrow when the growing season progresses.

Q2: The results are interesting because in this case the scenario analysis converges better towards the true value than the regression analysis. However, if good regression models can be obtained (like in Morocco), than it is usually more reliable to use the regression analysis over the scenario analysis because its error bounds are narrowing during the season and the analysis itself has a better theoretical foundation compared to the scenario analysis.