BACK-TO-OFFICE REPORT

ROME/MANILA, 20120910-15

TRAINING ON CROP YIELD MODELLING IN THE FRAME OF THE AMICAF PROJECT

René Gommes, Agrometeorologist consultant
rene.gommes@agro-impact.com

20120921, rev. 20121003

1. The reporter gave two days of training with exercises on the simulation of crop yields using the WABAL component of MOSAICC. Altogether, MOSAICC performed excellently, in particular the integration of the modules[1]: the WABAL component accessed without any problem grids of rainfall that had been prepared in a previous training session using the AURELHY component. Average province WABAL outputs were computed from the grid upon simple provision of the province Shapefile. It is suggested that MOSAICC could be used in non climate-change applications, such as crop monitoring, crop forecasting, index-based insurance etc.

2. The training material (PowerPoint presentations in pdf format) was made available to project staff.

3. Production wise, the major Philippine crops are sugar-cane, rice, coconut, bananas and maize (see figure 1). The training covered only rice and maize, but the five major crops could be covered at little additional cost. Sugar-cane should not pose any specific difficulty with WABAL, but coconut would require some additional exploratory work, mainly because the timing from flower initiation to harvest typically covers three years in palms. As to bananas, the major difficulty may be linked with available statistics which tend to be unreliable with most crops with indefinite phenology, such as bananas and cassava.

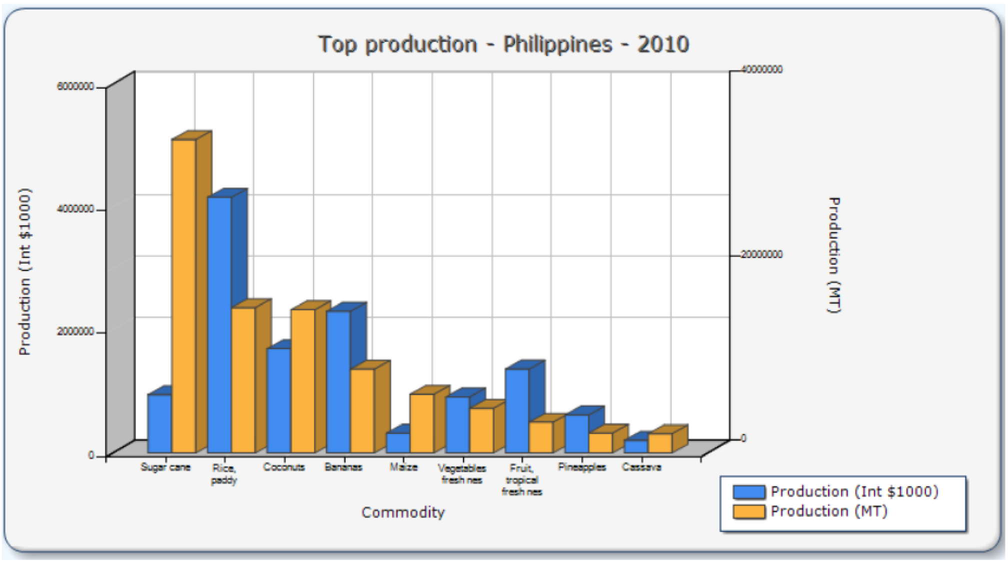


Figure 1: Main crops produced in the Philippines, ranked by production amount. Source: FAOSTAT.

1 The reporter also noted the excellent remote & real-time support provided by the developer of the MOSAICC shell, which indicates a well-oiled Rome-Pisa-Manila team approach!

4. The project intends to limit the future yield projections to the areas of main production, which accounting for about 90% of production. This makes little sense, for several reasons:
   (a) it is not known which crops will be grown where in the future. It may be, that the potential will change, with resulting shifts in national cropping patterns. In itself, this is a piece of information which may be useful to know and which can be derived using MOSAICC, but only if the whole country is covered;
   (b) the additional effort, again, will be minimal, and showing maps without spatial gaps is desirable for decision makers;
   (c) it is difficult to justify ignoring 10 % of production in a country where the inter-annual variability is lower than 10%;
   (d) it may well be that the inter-annual variability at the national level is brought about by the production in the more marginal areas that make up the gap between 90% and 100% of production.

5. The rule of all yield-weather modelling is that use must be made of whatever weather and crop data are available. Data for about 40 synoptic stations were obtained from PAGASA. While the network could be expanded with additional data from non-PAGASA sources in strategically important areas, it is suggested that the coverage is satisfactory, overall[2]. Nevertheless, some work should be done on radiation estimation based on maximum temperature along the lines of the Hargreaves-Samani approach (reference provided), calibrated against existing radiation stations and Campbell-Stokes sunshine durations. Once ten-day radiation data have been estimated, they need to be spatially interpolated. Additional menu items could be included in MOSAICC to include some "raw" climate variables by phenological phases (sample provided in Annex I), to serve as additional explanatory variables[3]
   (a) solar radiation by crop phase (defined by, in addition to the WABAL water balance variables (ETA, Deficit, surplus);
   (b) extreme maximum temperature, as this is a crucial qualitative factor affecting male sterility in rice and, indirectly, yield as well.

6. Yield calibration by single provinces is unlikely to yield acceptable yield functions, nor is this desirable because of the large number of functions that will need to be generated. The preliminary tests carried out for the training were promising (e.g. for white maize and palay, illustrated in fig. 2), but **they** also **showed rather optimistic correlation intensities due to the limited number of data points**. It is suggested to (1) mix cross-sectional data (neighbouring provinces) (2) time series data and (3) different harvesting dates to achieve higher statistical confidence of the yield functions. It is also suggested that the above-mentioned radiation data and extreme maximum temperatures may be necessary in some areas to reach acceptable yield functions.

---

2 i.e. the additional gain in precision of grids will be pointless in view of the deficiencies of other data, i.e. agricultural statistics.

3 This is easily done based on the values of F1, F2 and F3:

`if (dek-pld)/cycle<=f1` crop is is phase 1 (initial)
`if (dek-pld)/cycle>f1 and (dek-pld)/cycle<=f2` crop is in phase 2 (vegetative)
`if (dek-pld)/cycle>f2 and (dek-pld)/cycle<=f3` crop is in flowering stage
`if (dek-pld)/cycle>f3` crop is maturing (phase 4)

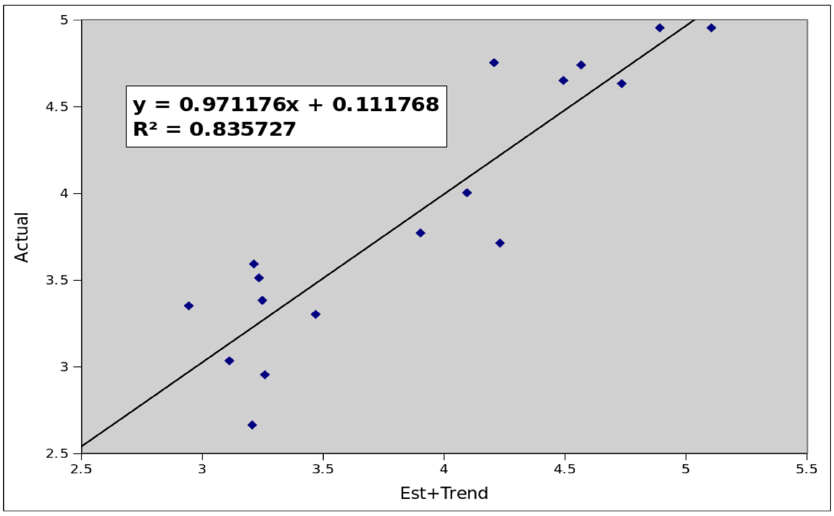


Figure 2: Actual Q4 (4th quarter) palay (paddy) yield in Nueva Ecija between 1994 and 2010 (17 years) by adoption of a linear trend (R^2=0.37) and the water balance variables ETA2, ETA3, ETA4, EXC4 and TWR. Although the R^2=0.835 is quite acceptable, it is nevertheless based on a large number of explanatory variables, i.e. 5 water balance variables and two trend variables.

7. The derivation of yield functions is a complex exercise, and it was not possible to enter any level of detail in the training. The following points were mentioned but may **need exploring in practice by project staff to be properly understood**:
   (a) statistical significance vs. agronomic significance in general, and in particular as regards (i) the sign of regression coefficients, (ii) the length of the calibration period and (iii) the type of technology trend (essentially logistic vs. linear[4]). Refer to Annex II for two illustrations of logistic trends which model actual yields much better than linear trends;
   (b) the avoidance of multi-collinearity in explanatory variables and the level of significance of regression coefficients (illustrated by manually selecting variables with a ratio of coefficient/standard error > 2);
   (c) the assessment of the "stability" of the equations by using jackknife-like techniques, separate calibration of first-half vs. second half of data set, random grouping of data etc.
   (d) the assessment of the effect of spatial aggregation, in particular whenever no usable equations can be achieved at a low level of aggregation (province of group of provinces);
   (e) finally: it would be useful is calibration equations (coefficients) could be "fed" to MOSAICC, as this would greatly simplify future yield projections.

   It is recommended that project staff responsible for the calibration of yield functions seek the advice of a qualified statistician when embarking on the determination of the yield functions. The statistician should help with the identification of techniques to select variables and the identification of regression software that can carry out supervised addition and deletion of variables.

4 An excellent tool to carry out non-linear detrending is Curve Expert Professional (CEPro) , but there are, of course, others... For CEPro, see http://www.curveexpert.net/products/curveexpert-professional/

ANNEX I

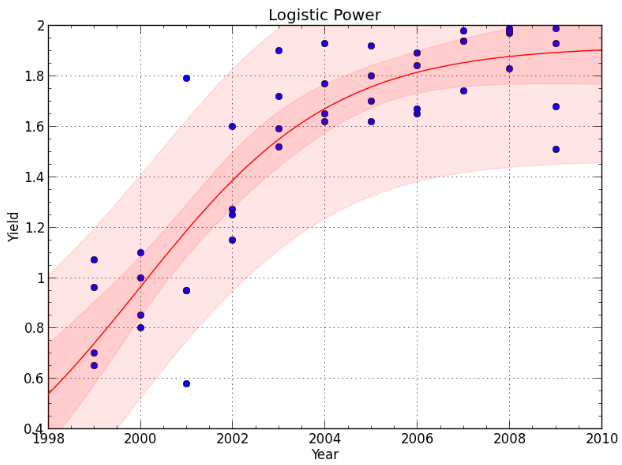
Sample WABAL output presenting, next to the standard MOSAICC/WABAL variables, radiation and extreme daily maximum temperature per dekad in Muñoz research station. The crop is assumed to be maize with the following parameters: cycle=12 dekads, pre-season dekads=5, pre-season crop coefficient 0.5, water holding capacity=100mm, no irrigation, f1=0.2, f2=0.5, f3=0.8, k1=0.5, k2=1.15 and k3=0.6. Crop planted at dekad 25.

year	eta-t	eta-i	eta-v	eta-f	eta-m	exc-t	exc-i	exc-v	exc-f	exc-m	def-t	def-i	def-v	def-f	def-m	TWR	wsi	wsicorr	rad_i	rad_v	rad_f	rad_m	txx_i	txx_v	txx_f	txx_m
1991	290.7	40.6	144.9	105.1	0.0	356.8	285.5	71.3	0.0	0.0	595.1	0.0	34.3	227.8	332.9	410	71	71	51.7	70.7	48.2	37.6	33.0	32.5	32.0	33.0
1992	280.5	38.5	149.4	92.5	0.0	150.5	150.5	0.0	0.0	0.0	588.0	0.0	37.4	218.0	332.5	409	69	69	62.9	77.5	61.6	37.8	33.5	34.0	34.4	33.0
1993	352.2	40.7	143.9	130.0	37.3	322.0	204.8	117.2	0.0	0.0	451.7	0.0	0.0	168.4	283.3	400	88	88	55.1	70.0	48.9	33.0	34.0	34.0	34.0	33.0
1994	257.9	42.5	158.9	43.1	13.3	169.9	133.9	36.0	0.0	0.0	714.6	0.0	43.7	345.0	325.9	449	57	57	53.2	66.4	51.5	32.9	33.8	35.1	34.6	34.4
1995	376.1	34.9	150.4	123.6	67.2	435.5	237.3	151.7	46.4	0.0	294.9	0.0	4.8	75.9	214.1	387	97	97	76.0	103.6	76.1	49.3	33.5	34.5	33.0	31.5
1996	358.8	44.1	154.4	125.3	34.8	202.6	144.2	40.1	18.3	0.0	354.5	0.0	0.1	58.7	295.6	414	87	87	60.1	73.6	50.5	38.2	33.0	33.5	33.5	33.5
1997	180.7	45.7	133.8	0.2	0.9	78.9	45.8	33.1	0.0	0.0	885.9	0.0	175.9	381.5	328.4	439	41	41	72.6	91.2	65.9	41.0	33.5	34.0	34.0	32.5
1998	385.5	39.3	141.4	125.8	78.9	418.0	267.5	150.5	0.0	0.0	250.2	0.0	9.9	136.7	103.5	386	100	100	62.4	79.9	53.7	35.4	33.5	33.7	33.0	32.5
1999	332.8	39.7	149.5	109.3	34.2	298.1	209.2	88.9	0.0	0.0	533.3	0.0	29.7	218.2	285.4	400	83	83	49.9	72.2	52.9	33.4	32.5	34.0	33.5	33.0
2000	354.7	41.1	146.4	125.5	41.6	398.8	83.6	315.2	0.0	0.0	379.8	0.0	0.0	100.2	279.6	397	89	89	58.8	65.0	46.6	36.5	34.7	34.2	33.8	33.7
2001	220.0	44.1	136.0	15.1	24.6	61.8	61.8	0.0	0.0	0.0	855.6	0.0	190.2	363.0	302.3	434	51	51	54.2	80.8	55.4	38.8	34.2	35.4	34.0	32.5
2002	307.4	43.2	160.1	104.0	0.0	359.4	219.1	140.3	0.0	0.0	543.1	0.0	0.0	210.3	332.8	437	70	70	62.9	85.0	66.6	40.8	34.2	35.5	35.0	34.0
2003	312.9	45.2	172.3	95.4	0.0	312.3	209.1	103.1	0.0	0.0	588.5	0.0	35.4	226.0	327.1	439	71	71	67.6	88.7	68.3	47.0	34.5	35.5	34.5	34.5
2004	388.1	50.7	166.7	80.1	90.5	212.7	113.0	40.8	58.8	0.0	447.3	0.0	112.4	198.8	136.1	446	87	87	60.0	74.5	45.3	38.4	35.0	35.0	33.0	32.5
2005	399.0	43.7	160.3	134.4	60.3	387.0	132.9	254.1	0.0	0.0	410.1	0.0	0.0	144.9	265.1	427	93	93	51.8	79.1	48.7	36.0	34.5	34.5	33.0	32.5
2006	420.6	46.7	157.0	138.6	78.1	326.0	128.8	157.5	39.6	0.0	275.3	0.0	0.0	73.2	202.0	432	97	97	61.3	79.3	57.1	37.6	33.5	33.5	34.5	32.5
2007	424.8	46.5	160.7	129.0	88.6	314.5	101.6	190.7	22.1	0.0	235.0	0.0	77.9	58.8	98.2	425	100	100	61.9	78.9	56.1	38.4	34.0	35.5	33.0	33.5
2008	325.4	48.8	165.0	111.4	0.0	221.1	75.1	146.0	0.0	0.0	619.8	0.0	62.9	226.2	330.6	440	74	74	56.1	73.8	55.4	35.1	34.5	35.0	34.0	33.0
2009	267.2	43.2	156.3	67.6	0.0	574.6	230.8	343.7	0.0	0.0	663.7	0.0	22.3	303.5	337.9	448	60	60	52.1	69.9	56.3	36.3	34.0	34.0	34.5	32.5

ANNEX II

Non-linear corn trends in Nueva Ecija province as assessed with Curve Expert Pro. Stagnating yields over recent years are clearly visible and constitute a strong argument against linear trends.

White corn



Family: Sigmoidal Models (logistic power)

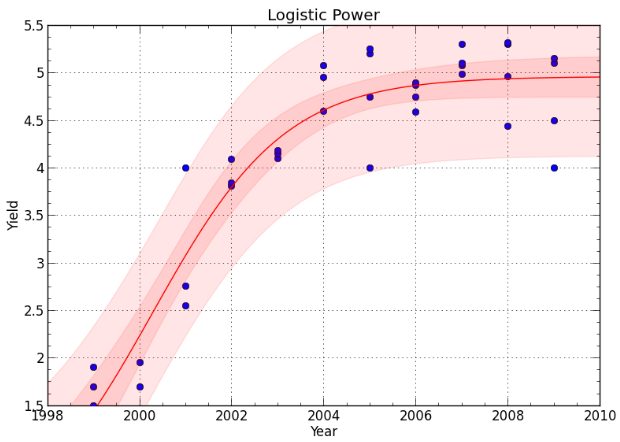
Equation: a/(1+(x/b)**c)

Parameters:
a = 1.92148669858072E+00
b = 1.99997314657660E+03
c = -9.49176861027706E+02

R^2= 7.72089E-01
Linear R^2 = 6.69955E-01

Parameter Standard Deviations:
a_stddev = 8.02780817600445E-02
b_stddev = 3.03612509819052E-01
c_stddev = 1.92390896168633E+02

Yellow corn



Family: Sigmoidal Models (logistic power)

Equation: a/(1+(x/b)**c)

Parameters:
a = 4.96804984278371E+00
b = 2.00027612451398E+03
c = -1.38401383940521E+03

R^2= 8.89993E-01
Linear R^2 = 6.80888E-01

Parameter Standard Deviations:
a_stddev = 1.07320940956202E-01
b_stddev = 1.77294671217429E-01
c_stddev = 1.87298476772981E+02