Integrated assessment of business crop productivity and profitability to use in food supply forecasting

Bojar W., Knopik L., Żarski J.

Abstract— Climate change suggests long periods without rainfall will occur in the future quite often. The previous approach showed that the dependence of crop-yields due to the size of rain is statistically significant. The paper examines the proposals to build a model describing the amount of precipitation and taking into account periods without rain. This model is based on a mixture of gamma distribution and one point-distribution. In the work on the basis of rainfall data, one can estimate parameters of the mixture. Estimators of these parameters were constructed using the method of maximum likelihood. Available data on precipitation allows for the confirmation of the hypothesis of the adequacy of the proposed model's rainfall. Long series of days or decades without rainfall allow one to determine the probabilities of adverse developments in agriculture (droughts). This could be the basis for forecasting crop yields in the future. Forecasted crop yields in the future (the years 2030 and 2050) can be used for assessment of the productivity and profitability of some selected crops in Kujavian-Pomeranian region. Assumptions and parameters of large-scale spatial economic models will be applied to build up relevant solutions. Calculated with this approach, output could be useful in expecting a decrease in the agricultural output in the region. It will enable one to shape effective agricultural policy in order to know how to balance the food supply and demand through appropriate managing with stored raw food materials and/or import/export policies. Used precipitation-yields the dependencies method allowing one to verify the earlier used methodology through a comparison of the obtained solutions concerning forecasted yields and close to it to an uncertainty analysis.

Index terms — climate changes, droughts' probability occurrence, expected yields, forecasted outputs

1 Introduction

Precipitation is a key component that links the atmosphere and the portions of the Earth's surface where water is in solid form through complex processes. Thus, accurate knowledge of precipitation levels is a fundamental requirement for improving the prediction of weather systems and of climate change. In particular, some of the weather forecast models use data assimilation techniques that require accurate precipitation estimates at high spatial and temporal resolutions.

In many applications, such as climate change, hydrology and meteorology, rainfall phenomena play a significant role. In numerous applications, the processes involved depend on microstructure of rain. The microstructure of rain is defined by raindrop size distribution, which represents the expected number of the raindrop per unit of the raindrop's diameter interval and the per unit volume of air.

This is in the most popular form, although numerous studies assume more general exponential or gamma distributions. The problem with the approach is that the raindrop size distribution definition assumes the temporal stationary and spatial homogeneity of the rain, which is in practice never reached, even at a small scale of the order of typical inter-drop distance.

Observed climate change causes undesired trends in the precipitation level and frequency of what it can influence essentially yields and agricultural production output.

On the basis of the proved precipitation changes in the Kujavian-Pomeranian regions in Poland, one can forecast crop yields in the future (for the years 2030 and 2050) [Bojar et al. 2013]. It can be used for the assessment of productivity and the profitability of some selected crops the in Kujavian-Pomeranian region. Assumptions and parameters of large-scale spatial economic models will be applied to build up relevant solutions. Socio-economic and technical criteria could be useful in the study of and especially forecasted cropping areas, yields, prices and costs, which will be employed to estimate the possible economic effects of forecasted more often periods of droughts.

Calculated with this approach, the output could be useful for anticipating a decrease in the agricultural output of the surveyed crop (spring barley) yield in the region. It will enable one to shape, in a wider scope, effective agricultural policy to know how to balance the food supply and demand through the appropriate management of stored food raw materials and/or import/export policies.

Secondary, met with the abovementioned methods, precipitation-yield dependencies allow one to verify an earlier-used methodology [Bojar, Knopik 2013] through a comparison of obtained solutions concerning forecasted yields in the future and closed to it an uncertainty analysis.

2 Rainfall distribution model

We consider that the continuous random variable T describes the value (size) of rainfall per day. We assume that T has a cumulative distribution function is F(t, α, β) with F(0, α, β) = 0. In this paper, we assume that F is a two-parameter gamma distribution with parameters (α, β). The density function is the following:

$$f(t:\alpha,\beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} t^{\alpha-1} e^{-t/\beta} \qquad \text{where } t > 0,\ \alpha > 0,\ \beta > 0. \tag{1}$$

Γ(α) describe the gamma function given by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

We assume that the days without rainfall are recorded as '0' so that the modified distribution has the density function given by:

$$g(t:p,\alpha,\beta) = \begin{cases} 1-p & \text{for} \quad t=0 \\ pf(t:\alpha,\beta) & \text{for} \quad t>0 \end{cases} \tag{2}$$

The mean value of the random variable T is ET = pαβ.

The purpose of this chapter is to consider the probability density function given by (2), when F is a two-parameter gamma distribution. The problem of statistical inference about (p, α, β) has received considerable attention [Kale 2000, Kale, Muralidharan 2000, Muralidharan, Kale 2002]. In [Muralidharan, Kale 2002] considered the case where F is a two-parameter gamma distribution with the shape parameter β and scale parameter α, and they obtained a confidence interval for α and β. In the paper [Muralidharan, Kale 2002], this distribution is considered and the maximum likelihood estimation of parameters (p, α, β) is obtained. In [Muralidharan, Kale 2002], the approximate $1-\gamma$ confidence interval for the mean value of T is the following:

$$(p\alpha\beta - u_\gamma \frac{\sqrt{K}}{\sqrt{n}}, p\alpha\beta + u_\gamma \frac{\sqrt{K}}{\sqrt{n}}) \tag{3}$$

where

$$K = p\alpha\beta^2[1+(1-p)\alpha]$$

and u_γ is the value of the standard Gauss random variable U such that

$$P\{|U| < u_\gamma\} = 1-\gamma \ .$$

2.1 Application of the model

Case A. We analyze the data set that contains the value of daily rainfall in decades.

The decades observed from April 1st to September 30th during the years 1999 to 2012. Statistical analysis was done for n = 252 decades. The probability distribution of rainfall shows good coordination with the distribution (2). The estimated parameters have the values p = 0.071, α = 1.2 and β = 17.27.

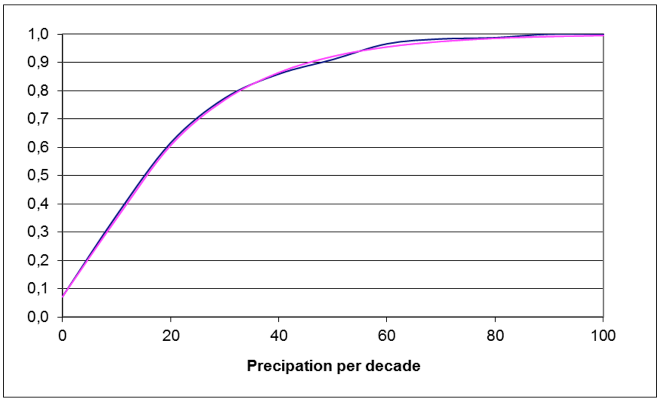

Fig. 1. Empirical and theoretical distribution functions

A concordance of the empirical distribution and the theoretical distribution test was carried out using the classical λ – Kolmogorov test and the $\chi2$ – Pearson test. The value of the test statistics $\lambda = 0.17$. For the goodness of the fit test, the $\chi2$ – Pearson statistic calculated $\chi2 = 7.92$, p – value = 0.44. This confirms concordance of the model with empirical data. Figure 1 shows the distribution of the empirical and theoretical distribution functions. The average value calculated from the data is ETe = 20.27 and the standard deviation s = 18.21. The average value calculated from the model is ET = 20.73, while the standard deviation DT = 18.92. The confidence interval for the mean value, with a confidence level of $1 - \gamma = 0.95$, was determined from the formula (3) and is as follows: (19.36, 21.17).

Case B. In this case, the precipitation is analyzed due to the day. There are also data from 1999 – 2012 and the months spanning June 1st to August 20th. Together, there was an analysis of n = 368 data on rainfall. The values of the estimated parameters are p = 0.61, $\alpha = 0.79$ and $\beta = 8.18$.

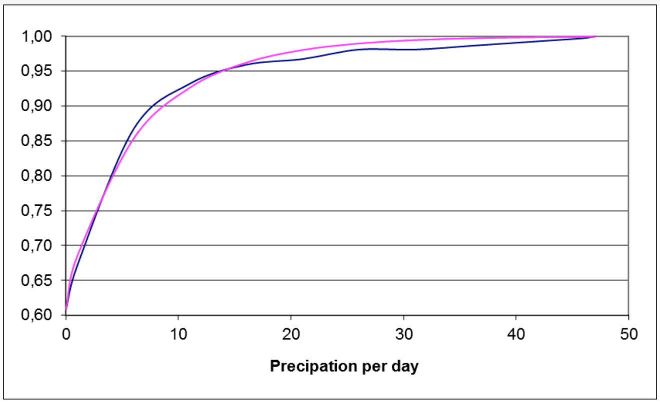

Fig. 2. Empirical and theoretical distribution functions for daily rainfall

A concordance of the empirical distribution and the theoretical distribution test was carried out using the classical λ – Kolmogorov test. The value of the test statistics is $\lambda = 0.26$. For the goodness of the fit test, the $\chi2$ – Pearson statistic is $\chi2 = 7.92$, and p – value = 0.44. This confirms the concordance of the model with empirical data. Figure 2 shows the distribution functions as empirical and theoretical. The average value calculated from the data are ETe = 7.16 and the standard deviation is s = 9.34. The mean value is calculated from a model ET = 6.45, while the standard deviation DT = 7.26. The confidence interval for the mean value with a confidence level $1 - \gamma = 0.95$ is determined from the formula (3) as follows: (6.33, 7.98).

2.2 Prognosis of long-series decades without precipitation

On the basis of the probability distribution model presented above, we predicted of a long series of decades where there were days without rain. The application of the Monte Carlo method generated 18 of the elements series of decades (six months). On this basis, the likelihood determined decades without rain as a series of a given length k (k = 7, 8, 9). If p_k is the probability of the occurrence of at least one k – element series without rain, then $(1 - p_k)^m$ is the probability of the absence of the series k – element with no rain in the next m years. Let p_{km} be denoted by the probability of at least one series of k – element in the following m – years without rain. For the assessment of the occurrence of a long series of decades without rain until 2030 by assuming m = 17, and 2050 m = 27. Table 1 below shows the values of the probability p_{km} for the years 2030 and 2050 and the series length k = 7, 8, 9, 10.

Table 1. Probability of the occurrence of a series of decades without precipitation until 2030 and 2050

Year	Series length	Probability
2030	7	0.199
	8	0.085
	9	0.035
	10	0.010
2050	7	0.393
	8	0.180
	9	0.076
	10	0.021

The calculation results contained in Table 1 show that, for the length of the series decades without rain equal to k = 7, 8, the calculated probability is relatively high. This means that the probability of the occurrence of extreme weather events in the considered span of time is high.

The analysis of statistical data on yields shows that low cereal yields are associated with long periods without precipitation. This fact will be the basis for building a simple regression model enabling one to predict the occurrence of adverse events in the cultivation of selected plants. Below, the occurrence of a significant dependence of the yield of barley on the length of successive decades without precipitation is shown. On the basis of data on the yields of barley and the length of the series of decades without rain, a regression line describing the dependence of the yield of barley on the number of decades without rain was determined. A regression line takes the form of yield = length of series * (–1.544) + 34.8995. The correlation coefficient is R = 0.595; the square of the correlation coefficient is called the 'coefficient of determination' and defines the percentage of variation explained by the equation, R2 = 0.354. Testing the significance of the regression equation was performed using the F test; the calculated value of the F – statistics = 6.58 and the significance level corresponding to this value p – value = 0.0247. On this basis, it is concluded that the proposed regression equation is statistically significant.

Table 2. Yield of spring barley for different length of series

Length of series	Yield (tonnes/ha)
7	2.409
8	2.255
9	2.100
10	1.946

The findings described above can be applied for forecasting yields of spring barley in the Kujavian and Pomeranian region.

For 2030 and for 8 series of decades without rain at a probability of 8.5% one can forecast a spring barley yield at levels equal to 2.255 tones/ha while for 2050 and for 8 series of decades without rain at a probability of 18.00%, one can expect the same level of a spring barley yield equal to 2.255 tonnes/ha (see Table 1 and Table 2).

3 Economic approach for forecasting regional output of agricultural production based on spring barley cropping

In regional case study analysis, data from CAPRI (large-scale model) model [Britz, Witzke 2012, Stocco at al. 2013, Köchy, Zimmermann. 2013] were used based on the following assumptions. For the survey data regional resolution, NUTS2 was considered. K&P belongs to the Northern Region (NUTS 1), PL61 voivodship (NUTS 2), while Lubelskie belongs to the PL3 Eastern Region (NUTS 1), PL31 voivodship (NUTS 2).

The SSP2 socio-economic adaptation and mitigation challenges (fossil and recourse intensity) are at a medium level and the present climate within AgMIP S1 is assumed. SSP2 is called "Continuation".

To characterize the outputs of some important crops within the K&P region, some calculations based on CAPRI databases and the agro-climatic findings from the UTP and the IAL PAS were made. They concerned farmland areas, yields and input prices from 2010, 2030 and 2050 for modelling crops.

The obtained findings (see Table 1) show that extreme lengths of time without precipitation occur for the series number 8 decade with a high probability in 2030 and 2050, which determines a high risk of the occurrence of such extreme natural events.

The total volume of output of the selected crop of the region expressed in physical units is calculated according to the formula (4):

TRO = L*Y where: (4)[1]

TRO- total regional output of a given crop (a number of thousand tonnes)

L –land area of a given crop (number of hectares)

Y – yield of a given crop (number of tonnes per 1 hectare) (from the CAPRI database or from the UTP findings)

Next, values of the production of particular crops within the region were calculated to compare the findings based on the CAPRI model and the findings based at linked CAPRI and the agro-climatic models of the University of Technology and Life Sciences according to the formula (5):

TOV = L*Y *P where: (5)[2]

TOV – total regional value of output of a given crop (per thousand euro)

L – land area of a given crop (number of hectares)

Y – yield of a given crop (number of tonnes per 1 hectare) (from the CAPRI database or from the UTP findings)

P – producer prices (euro/t)

Table 3. The regional data setup on a wide area model and the UTP findings

	Producer Price [euro/t}			Land [1000 ha or hds]			Average yield [t/ha]		
CROP	2010	2030	2050	2010	2030	2050	2010 (CAPRI)	2030 (UTP)	2050 (UTP)
Spring barley	93.07	128.89	166.21	111.45	89.24	91.30	3.432	2.255	2.255

Explanation: All data concerning prices and land for the Kujavian and Pomeranian region is based on the CAPRI database while yields of spring barley are based for 2010 on the CAPRI database when 2030 and 2050 are based on the UTP findings.

The facts expressed in Figure 3 and Figure 4 allow one to see that the forecasted extreme weather

[1] All parameters in this formula were set up using methods described within the CAPRI database.
[2] All parameters in this formula were set up using methods described within the CAPRI database.

conditions strongly affect the agricultural output of selected crops in the Kujavian and Pomeranian region, which are essentially different from the agricultural outputs set up on modelling assumptions (for average yields, see Table 3).

Hence, specific detailed analysis shows the undesirable drops in the frequency of precipitation can influence essentially agricultural outputs calculated on the assumed average yields in the region in the future, which increases the probability of this occurrence in the region and at the same time it can cause an imbalance in the food supply and demand.

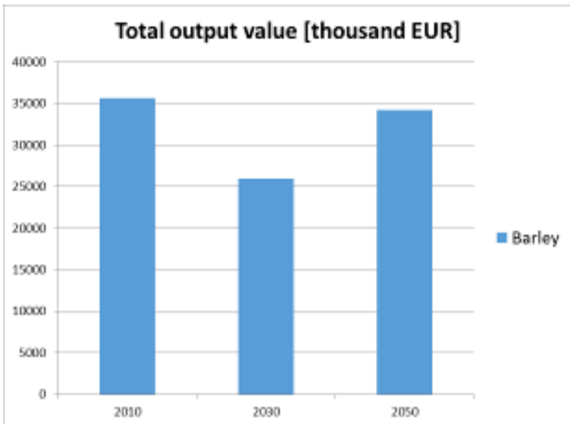

Fig. 3. Comparison between CAPRI K&P and CAPRI-UTP model value output (for average yield)
Source: own study based on CAPRI and Agro Climatic UTP data and models

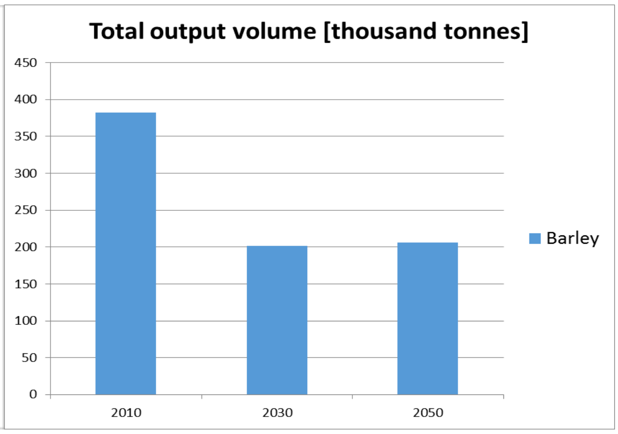

Fig. 4. Comparison between CAPRI K&P and CAPRI-UTP model volume output (for average yield)
Source: own study based on CAPRI and Agro Climatic UTP data and models

4 Conclusion

The calculation of the future agricultural output volume and value in the Kujavian and Pomeranian region affected by the changes in yields of spring barley allows for an estimate associated with it and a risk of an imbalance in the food supply and demand. The found solution is based on expected changes to barley's yield due to precipitation and its distribution, extreme changes in 2030 and 2050 and large-scale spatial model assumptions described briefly in Ch. 4. This allows one to create some forecasts with a defined probability of the occurrence of extreme output changes compared to average ones. Projections of producer prices from the selected model baseline scenarios (GAMP) were also possible after the comparison of models based on regional empirical data. This allows one to forecast the levels of outputs of selected crops in agriculture in the surveyed regions in 2030 and 2050, calculated according to differentiated simulated assumptions. It can help with conducting a more appropriate agricultural and trade food policy to ensure food safety in different spatial scales of Europe and balance the food supply and demand in the perspective of the next 40-50 years.

Presented in this elaboration, the method of forecasting long series without precipitation was positively verified through a comparison of the findings from the application made with other methods when the forecasted yield of spring barley was comparable to the yield calculated in Table 2 at comparable level of the rate of probability of occurrence [Bojar, Knopik 2013].

References

Bojar W., Knopik L., Żarski J., 2013. Analiza wpływu warunków klimatycznych na plonowanie roślin uprawnych w regionie kujawsko-pomorskim (Analysis of the impact of climate conditions on crop yields in the Kujavian-Pomernian region). Studies & Proceedings of the Polish Association for Knowledge Management. No. 64. pp. 31-44.

Bojar W., Knopik L. 2013. Paper presented at MACSUR project workshop at the Leibniz Centre for Agricultural Landscape Research (ZALF), Mücheberg, Germany November 18-20, 2013 titled: "Integrating TradeM and CropM MASCUR models for regional case studies in Poland". (University Bydgoszcz, Poland) http://ocs.macsur.eu/public/conferences/2/schedConfs/2/program-en_US.pdf)

Britz W., Witzke P. 2012. CAPRI model documentation Editors1: http://www.capri-model.org/docs/capri_documentation.pdf.

Kale B. K., 1999. A First Course on Parametric Inference, Narosa Publishers, New Deli.

Kale B. K., Muralidharan K., 2000. Optimal estimating equations in mixture distributions accommodating instantaneous or early failures, Journal of Indian Statistical Associations, vol. 38, pp. 317–329.

Köchy, M., A. Zimmermann. 2013. Workshop on regional pilot studies, June 5–7, 2013, Braunschweig. FACCE MACSUR Reports 2(1): H1.2-R1.

Muralidharan K., Kale B.K., 2002. Modified gamma distribution with singularity at zero, Communications Statistics, Simulation Computer, vol. 31, No1, pp. 143 – 158.

Stocco, L., Adenäuer, M., Zimmermann, A. (2013): Global land use response in agricultural sector models: estimating the supply and area response in Argentina. Selected paper, 133rd EAAE seminar "Developing Integrated and Reliable Modelling Tools for Agricultural and Environmental Policy Analysis", June 15-16, Chania, Crete, Greece.