Analyzing the value of the residue of crops grown on arable land with stochastic optimization

Ádám Zlatniczki

INFO
Received 17 March 2015
Accepted 16 Apr 2015
Available on-line 29 Jun 2015
Responsible Editor: M. Herdon

Keywords:
value of crop residue, optimal artificial fertilizer allocation, stochastic optimization, Monte Carlo simulation, linear programming

ABSTRACT
Determining the value of the residues of crops grown on arable land is a non-trivial task, depending much on how it is defined. In this paper the value of residues is considered to be the savings achieved on the total expense of artificial fertilizer distribution by returning the residues into the soil. A general linear programming approach is presented to obtain optimal artificial fertilizer allocation. Since the amount of artificial fertilizers required depends on uncertain inputs Monte Carlo simulation is applied in conjunction with linear programming to solve the arising stochastic optimization problem. The input data, such as the national average yield, specific amounts of nutrients required by crops to achieve the national average yield, and publicly available details of different artificial fertilizer products are specific to Hungary, but the mathematical model presented is general in nature. Simulations are executed for some of the major crops, including wheat, corn, sunflower and rape. The distribution of savings achieved on returning their residues to the soil is provided at the end of the paper for further use.

1. Introduction

The value of crop residues can be defined in many ways, and depending on this definition different qualitative or quantitative measures can be determined. In this paper the definition is chosen to be the savings achieved on artificial fertilizing caused by the amount of nutrients returned into the soil, which can be described with a quantitative measure, namely the difference of the total expense of the fertilizing process with and without returning crop residues to the soil. Determining the value of crop residues this way used to be an elaborate and complex task, but its results are extremely useful – especially in financial decision support. It is required to answer common questions such as whether crop residue should be collected and sold or returned to the soil instead.

The planning of artificial fertilizer allocation of different fields is a six-step process. First, the expected yield (in kg/m²) of the crop sowed into a field has to be decided. Next, the amount of nutrients – nitrogen (N), phosphorus (P) and potassium (K) – stored in the field are measured. The third step is to determine the specific nutrient requirements (kg of nutrient / kg of yield) of the crop. After this, the quantified nutrient requirements are calculated (in kg/m²). The fifth step is to perform the correction of the quantified nutrient requirements, and lastly the optimal amounts of the different artificial fertilizer products are allocated to the field (eds Nábrádi, Pupos & Takácsné, 2007).

The methods of operations research are scientific approaches to decision making, seeking to best design and operate systems under conditions, usually with the help of mathematical models. Such models consist of three components: objective function(s), decision variables and constraints. Optimization models seek to find the values of the decision variables that optimize (maximize or minimize) an objective function so that the given constraints are satisfied. If the objective function and the constraints can be written as linear combinations of the decision variables the optimization problem can be solved with linear programming (Winston & Goldberg, 2004). To this day many textbooks on agricultural management teach only heuristic approaches to solve such problems, including the allocation of artificial fertilizer. By applying these approaches one is unlikely to find an optimal solution - even though simple linear programs could be used instead to obtain optimal
allocations, they still lack public awareness in Hungary (even in the ranks of professional agricultural engineers). Since the 85-95% of the total expense of artificial fertilizing comes from the cost of the artificial fertilizer itself the optimal choice of products is unquestionably crucial (eds Nábrádi, Pupos & Takácsné, 2007).

There are important but difficultly quantifiable factors which have to be taken into consideration throughout the planning of artificial fertilizer allocation though. These factors are the amount of nutrients stored in the soil and the required amount of nutrients of a crop to achieve an expected yield. Their values are usually estimated based on lower and upper bounds provided in textbooks and professional experience. In a mathematical point of view they should be represented with random variables. Random variables take on specific values with specific probabilities – these value-probability pairs are described by probability distributions (Durrett, 2013). By including random variables we include uncertainty into the optimization model, thus arrive at a stochastic programming problem (King & Wallace, 2012). A common solution approach is simulation – we generate scenarios by taking samples from the probability distributions and solve the original problem for each scenario (Prékopa, 1995). One such sampling method is the Monte Carlo method – among all the numerical methods that rely on n-point evaluations in an m-dimensional space the absolute error of the Monte Carlo approximation decreases fastest, which gives the method an edge as the size of the problem increases (Fishman, 1996).

The application of linear programming or Monte Carlo simulation is not an extraneous concept in agricultural sciences either. Experimenting and prototyping are useful tools, but they are not able to provide on-time answers as opposed to simulation methods. A four-step iterative process is proposed in (Bergez, Colbach, Crespo, Garcia, Jeuffroy, Justes, Loyce, Munier-Jolain & Sadok, 2010) to design crop management systems with simulation. First a seed (consisting of strategies or decision rules) must be defined that is going to be used to generate crop management plans. In the second stage the simulation of such plans is executed. The third step involves evaluating the simulated plans, and lastly the interesting crop management options are selected and/or the seed is improved, starting a new iteration. Aldeseit used a linear programming model to determine least-cost synthetic fertilizer combinations and showed how important the application of linear programming might be (Aldeseit, 2014). A general mixed integer programming (MIP) model is introduced in (Hansson, Svensson, Hallefält & Diedrichs, 1999) that is capable of optimizing the amount of fertilizing products that have to be applied in each year of a cutting cycle of energy crops. Their model lacks the presence of uncertainty though, and it does not consider discretizing the total area into fields, which may affect optimal allocations. The model of Mínguez, Romero and Domingo approaches fertilizer allocation differently. Their model does not require nutrient requirements to be satisfied by all means – it views these requirements only as targets that should be achieved and penalizes the differences, allowing a more flexible and realistic specification of lower and upper limits of nutrients (Mínguez, Romero & Domingo, 1988). A possible application of simulation in conjunction with linear programming is shown to maximize the nutrient contents of compost manure prepared using pig dung, buffalo dung, green manure and concentrated super phosphate in (Minh, Ranamukhaarachchi & Jayasuriya, 2007). Simulation and linear programming are also applied in soil erosion control (Segarra, Kramer & Taylor, 1985), irrigation management (Li, Lu, He and Shi, 2014) and urban water management (Zhu, Marques & Lund, 2005), working schedule planning (Matsui, Inoue, Matsushita, Yamada, Yamamoto & Sumigama, 2005), organic farming risk management (Lauwers, Decock, Dewit & Wauters, 2010), etc.

2. Material and method

To model the total expense of fertilizing with and without returning crop residues to the soil Monte Carlo simulation is applied in conjunction with linear programming. First the mathematical model involved in Monte Carlo simulation is described in details, followed by the optimization model. Data is provided only for wheat, corn, sunflower and rape, but the mathematical models presented are independent from the data and can be extended easily.

2.1. Monte Carlo model
The Monte Carlo model is similar in nature to that of the four-step process proposed in (Bergez, Colbach, Crespo, Garcia, Jeuffroy, Justes, Loyce, Munier-Jolain & Sadok, 2010). The seed used for generating scenarios is defined as follows. It is assumed that the hypothetical total area (denoted with $TA$) involved in the calculations consists of 10,000,000 square meters (1,000 hectares) – an area of such magnitude ensures statistically representative results. The savings are simulated for each crop independently - it is also assumed that in every simulation only the kind of crop that is under simulation has been harvested from the whole area. The crops involved in the simulation are presented as the set $C$, such that

$$C = \{\text{wheat, corn, sunflower, rape}\}.$$  

The total area is then divided into $n = |C\setminus HC|$ fields – one field for the next generation of each different crop other than the one previously harvested (denoted with $HC$), in accordance with the crop rotation principle. To obtain general results the area of the fields (denoted with $A_i$) should vary between 0 m$^2$ and 10,000,000 m$^2$, but the sum of their expected values should equal the total area. Such partitioning can be achieved by using PERT distributions (Vose, 2008). The expected value of a PERT distribution is calculated as

$$E = \frac{a + 4 \times m + b}{6},$$

where $a$ and $b$ are the lower and upper bounds, and $m$ is the most likely value (Malcolm, Roseboom & Clark, 1959). The $a$ and $b$ parameters are 0 and $TA$ respectively, but the $m$ parameter has to be specified. To obtain PERT distributions with these lower and upper limits, and expected values summing to the total area we have

$$\frac{TA}{n} = \frac{0 + 4 \times m + TA}{6},$$

from which we get that

$$m = \frac{TA \times 6 - TA}{4}.$$  

Another constraint is that the sum of the areas of all the fields must equal the total area, which can be enforced as follows:

$$A_i \sim \min \left( \text{PERT}(0, m, TA), TA - \sum_{k=0}^{i-1} A_k \right), \text{where}$$

$$i \in \{1, 2, ..., n\} \text{ and } A_0 = 0.$$  

This constraint biases $A_2, A_3, ..., A_n$ though. To lessen the bias effects the values of $A_1, A_2, ..., A_n$ are permuted randomly in every iteration of the simulation. The permuted areas are then always assigned to the different crops in a linear order.

The yield of field $i$ can be defined as

$$Y_i = \int_{F_i} u(x, y) \, dF_i,$$

where $F_i$ represents the boundaries of field $i$, and $u(x, y)$ is the function describing the yield of field $i$ in its specific points. A simulation of fields as described in (5) does not specify the boundaries, only the area of each field. Even if the boundaries would be specified $u(x, y)$ would be a multivariate probability density function that had to be estimated. To simplify the problem fields are discretized into one dimension, namely into sequences of squared meters. This way the estimation problem boils
Ádám Zlatniczki: Analyzing the value of the residue of crops grown on arable land with stochastic optimization down to fitting univariate probability distributions to historical yield data. The lower and upper limits of the national average yield are provided in Table 1. The average yield, combined with the area of the previously described fields can be used to calculate the amount of nutrients returned by the crop residues to each field. The lower yield limits (LYL) and upper yield limits (UYL) are based on historical data from 2004 to 2014 stored by the Hungarian Central Statistical Office.

### Table 1. Lower and upper limits of national average yield

<table>
<thead>
<tr>
<th>Harvested Crop (HC)</th>
<th>Lower (LYL) [kg/m²]</th>
<th>Upper (UYL) [kg/m²]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat</td>
<td>0.4</td>
<td>0.6</td>
</tr>
<tr>
<td>Corn</td>
<td>0.4</td>
<td>0.8</td>
</tr>
<tr>
<td>Sunflower</td>
<td>0.2</td>
<td>0.26</td>
</tr>
<tr>
<td>Rape</td>
<td>0.2</td>
<td>0.3</td>
</tr>
</tbody>
</table>

It is acceptable to assume that the average yield of the different crops is uniformly distributed between these limits – the Kolmogorov-Smirnov p-values are 0.799, 0.1561, 0.799 and 0.871, respectively. That said, the yield of the kth square meter in field i (denoted with \(y_{i,k}\)) is described formally as

\[
y_{i,k} \sim U(LYLC_i, UYLC_i), \quad k \in \{1, 2, ..., A_i\},
\]

the HCi index denotes the crop that has been harvested from field i. The yield of field i (denoted with \(Y_i\)) is the sum of the yield of each square meter within said field. Since it is assumed that the yields of the square meters are independent and identically distributed, and the number of square meters in each field is expected to be very high, the yield of field i can be approximated with a normal distribution (in accordance with the central limit theorem) as follows:

\[
Y_i = \sum_{k=1}^{A_i} y_{i,k} \sim N \left( A_i \times E(y_{i,1}), \sqrt{A_i \times Var(y_{i,1})} \right) = N \left( A_i \times \frac{LYLC_i + UYLC_i}{2}, \frac{1}{12} \left( UYLC_i - LYLC_i \right)^2 \right),
\]

where \(E\) is the expected value operator and \(Var\) is the variance operator. The ratio of nutrients released from crop residue is provided in Table 2. These values are also essential in calculating the total amount of nutrients returned into each field.

### Table 2. Ratio of nutrients released from 1 kg of crop residue

(Sebestyén, Baranyai & Boldis 1983)

<table>
<thead>
<tr>
<th>Harvested Crop</th>
<th>Returned nutrient ratio (RNR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Wheat</td>
<td>0.005</td>
</tr>
<tr>
<td>Corn</td>
<td>0.006</td>
</tr>
<tr>
<td>Sunflower</td>
<td>0.008</td>
</tr>
<tr>
<td>Rape</td>
<td>0.004</td>
</tr>
</tbody>
</table>

The amount returned from nutrient j into field i is denoted with \(R^j_i\) and equals the product of the yield of field i (denoted with \(Y_i\)) and the returned ratio of nutrient j from the harvested crop on field i (denoted with \(RNR^j_{HC_i}\)), formally:

\[
R^j_i = Y_i \times RNR^j_{HC_i}, \quad j \in \{N, P, K\}. \tag{9}
\]
It is assumed that each generation of crop fully exhausts the soil, meaning that only $R_i^f$ is available for the next generation – the rest has to be supplied with artificial fertilizer.

The amount of nutrients required by a specific crop to achieve the national average yield varies between certain limits, as shown in Table 3.

Table 3. Nutrient amount ranges required for achieving the national average yield (eds Bocz, Késmárki, Kováts, Ruzsányi and Szabó, 1992)

<table>
<thead>
<tr>
<th>Planned Crop (PC)</th>
<th>Nutrient requirement [kg/m²]</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>P</td>
<td>K</td>
</tr>
<tr>
<td></td>
<td>Lower $LNL_i^N$</td>
<td>Upper $UNL_i^N$</td>
<td>Lower $LNL_i^P$</td>
</tr>
<tr>
<td>Wheat</td>
<td>0.0135</td>
<td>0.0135</td>
<td>0.0068</td>
</tr>
<tr>
<td>Corn</td>
<td>0.0120</td>
<td>0.0200</td>
<td>0.0066</td>
</tr>
<tr>
<td>Sunflower</td>
<td>0.0030</td>
<td>0.0080</td>
<td>0.0040</td>
</tr>
<tr>
<td>Rape</td>
<td>0.0050</td>
<td>0.0110</td>
<td>0.0070</td>
</tr>
</tbody>
</table>

Mathematically the amount required from nutrient $j$ in the $k$th square meter of field $i$ is a random variable, denoted with $X_{i,k}^j$. It is assumed that the values of these random variables are independently and identically uniformly distributed between the aforementioned lower and upper limits. The lower nutrient limit from nutrient $j$ for a square meter of field $i$ is denoted with $LNL_{PC_i}^j$, where the $PC_i$ index indicates the planned crop in field $i$ in next year’s crop rotation plan. $UNL_{PC_i}^j$ denotes the upper nutrient limit in a similar way. Formally,

$$X_{i,k}^j \sim U\left( LNL_{PC_i}^j, UNL_{PC_i}^j \right).$$

(10)

The nutrient requirement of field $i$ from nutrient $j$, denoted with $NR_i^j$, can be calculated by taking the sum of the nutrient requirement of every square meter in that field. Since it is assumed that every $X_{i,k}^j$ is independent and identically distributed the central limit theorem is applicable:

$$NR_i^j = \sum_{k=1}^{A_i} X_{i,k}^j \sim N \left( A_i \times E(X_{i,1}^j), A_i \times Var(X_{i,1}^j) \right) =$$

$$N \left( A_i \times \frac{LNL_{PC_i}^j + UNL_{PC_i}^j}{2}, A_i \times \frac{1}{12} \left( UNL_{PC_i}^j - LNL_{PC_i}^j \right)^2 \right).$$

(11)

At this point the seed for generating scenarios is fully defined – based on these rules the second step (the simulation of scenarios) can be executed. The third step of the process is the evaluation of each scenario. To get the value of the crop residues the total expense of fertilizing with and without returning the residues to the soil has to be determined - their difference is the actual amount saved on fertilizing. To obtain the total expenses the optimization problem described in the next chapter has to be solved.

### 2.2. Optimization problem

Calculating the total expense of fertilizing is a stochastic optimisation problem. The problem is solved by running Monte Carlo simulations to generate the possible nutrient requirement scenarios and applying linear programming to obtain the most cost efficient fertilizer allocation for each scenario. Table 4 contains publicly available data on artificial fertilizers, namely the ratio of nutrient $j$ in 1 kg of fertilizer $f$ (denoted with $RNF_{ij}^f$) and its price (with taxes). The products are indicated with the index $f \in \{AF1, AF2, ..., AF15\}$. The LP used is described as follows.
### Table 4. Publicly available data of artificial fertilizer products

<table>
<thead>
<tr>
<th>Product</th>
<th>Ratio of nutrients in 1 kg of fertilizer (RNF)</th>
<th>Avg. Price [HUF/kg]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>P</td>
</tr>
<tr>
<td>AF1</td>
<td>0.12</td>
<td>0.52</td>
</tr>
<tr>
<td>AF2</td>
<td>0.27</td>
<td>0.0</td>
</tr>
<tr>
<td>AF3</td>
<td>0.18</td>
<td>0.25</td>
</tr>
<tr>
<td>AF4</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>AF5</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>AF6</td>
<td>0.04</td>
<td>0.12</td>
</tr>
<tr>
<td>AF7</td>
<td>0.06</td>
<td>0.12</td>
</tr>
<tr>
<td>AF8</td>
<td>0.08</td>
<td>0.21</td>
</tr>
<tr>
<td>AF9</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>AF10</td>
<td>0.0</td>
<td>0.205</td>
</tr>
<tr>
<td>AF11</td>
<td>0.34</td>
<td>0.0</td>
</tr>
<tr>
<td>AF12</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>AF13</td>
<td>0.08</td>
<td>0.11</td>
</tr>
<tr>
<td>AF14</td>
<td>0.45</td>
<td>0.0</td>
</tr>
<tr>
<td>AF15</td>
<td>0.12</td>
<td>0.11</td>
</tr>
</tbody>
</table>


2.2.1. Variables

\( X_{i,f} \): Real-valued variables, representing the amount of artificial fertilizer \( f \) allocated to field \( i \) in kilograms (not to be confused with \( X_{i,k} \), which are random variables in the simulation).

\( Y_{i,f} \): The binary equivalent of \( X_{i,f} \) (a flag representing whether any amount of fertilizer \( f \) is allocated to field \( i \) – also not to be confused with \( Y_{i} \)).

\( Z_{i} \): The number of different types of artificial fertilizer allocated to field \( i \). \( Z_{i} \) is calculated as

\[
Z_{i} = \sum_{f} Y_{i,f}.
\]  

(12)

2.2.2. Objective function

The objective is to minimize the total expense of fertilizing. The expense of fertilizing consists of two components: the cost of artificial fertilizer and the aggregated cost of its distribution. The aggregated cost of distributing artificial fertilizer in 2014 in Hungary is 0.2889 Ft/m² (Gockler, 2014). It is assumed that only one kind of artificial fertilizer can be distributed simultaneously on each field – therefore the aggregated cost of distributing fertilizer on a field should be multiplied by the number of different fertilizers allocated to said field. The objective function can be written as

\[
\text{minimize} \sum_{i} \left( \sum_{f} X_{i,f} \times \text{Price}_{f} + Z_{i} \times A_{i} \times 0.2889 \right).
\]  

(13)

2.2.3. Constraints

The objective is subject to only two kinds of constraints, namely to satisfy the different nutrient requirements of each field and non-negativity. The right hand side of the nutrient requirement constraints depend on whether crop residues are returned to the soil or not. Equation (13.a) shows the constraints given no residue is returned, while equation (13.b) shows the (corrected) constraints when crop residues are returned. Equation (14) is the non-negativity constraint.

\[
\sum_{f} X_{i,f} \times \text{RNF}_{f}^{j} \geq \text{NR}_{i}^{j}, \forall i,j.
\]  

(14.a)
3. Simulation results

Reaching a convergent state is a key factor in case Monte Carlo simulations are involved. After 10,000 iterations the simulated results converge (Figures 1-4, convergent state represented by red dashed lines).

![Convergence of simulation (Wheat)](image1)

**Figure 1.** Convergence of wheat savings

![Convergence of simulation (Corn)](image2)

**Figure 2.** Convergence of corn savings

![Convergence of simulation (Sunflower)](image3)

**Figure 3.** Convergence of sunflower savings

![Convergence of simulation (Rape)](image4)

**Figure 4.** Convergence of rape savings

The expected value and standard deviation of savings are presented in Table 5. We can see that the standard deviations are very low, but considering that these values pertain only to a square meter the effects of uncertainty can vary in wide ranges given a high number of square meters involved.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Expected value [Ft/m²]</th>
<th>Standard deviation [Ft/m²]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat</td>
<td>2.4069174</td>
<td>0.0256387</td>
</tr>
<tr>
<td>Corn</td>
<td>2.7877342</td>
<td>0.0362423</td>
</tr>
<tr>
<td>Sunflower</td>
<td>0.9394069</td>
<td>0.0110463</td>
</tr>
<tr>
<td>Rape</td>
<td>4.0223666</td>
<td>0.1815932</td>
</tr>
</tbody>
</table>

To provide a better understanding of this uncertainty the probability distributions of savings are included as well (Figures 5-8, expected values represented by red lines).
Ádám Zlatniczki: Analyzing the value of the residue of crops grown on arable land with stochastic optimization

Figure 5. Relative frequencies of savings on wheat residue

Figure 6. Relative frequencies of savings on corn residue
Ádám Zlatniczki: Analyzing the value of the residue of crops grown on arable land with stochastic optimization

This uncertainty affects risks, thus has to be taken into consideration by financial decision making processes. Unfortunately no distributions fitted to the simulated data managed to achieve statistical significance based on the Kolmogorov-Smirnov, Anderson-Darling and $\chi^2$ goodness-of-fit tests. Instead, the minimum and maximum values, the quartiles, and the 5th and 95th percentiles are provided (Table 6).

Figure 7. Relative frequencies of savings on sunflower residue

Figure 8. Relative frequencies of savings on rape residue
4. Conclusions

A mathematical model has been created that utilizes Monte Carlo simulation and linear programming to obtain the distribution of savings on residues of crops grown on arable land. The model is based on general concepts and it can be used not only to determine the value of crop residues but also to optimize artificial fertilizer allocation for any number of fields with differing planned crops and residues. Although the model assumes that crops fully exhaust the soil nutrients left in the field can be included into the calculations by adding them to $R_i$. Application of this model has different positive effects as well. The permanent dosage of high amounts of artificial fertilizer sours the soil – by optimizing the required amount this process can be slowed down. The distributions of savings can also be involved in further simulations to support decision makers. It is important to note that these distributions are based on nationwide data – in different regions they might look different. Researchers or other individuals are welcome to the Python implementation of the model upon personal request or by downloading it from the journal web site.

References


Table 6. Quantiles of the savings distributions

<table>
<thead>
<tr>
<th>Residue</th>
<th>Min</th>
<th>$Q_{0.05}$</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.5}$</th>
<th>$Q_{0.75}$</th>
<th>$Q_{0.95}$</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat</td>
<td>2.3921713</td>
<td>2.3923930</td>
<td>2.3924884</td>
<td>2.3925728</td>
<td>2.4128814</td>
<td>2.4682789</td>
<td>2.5100094</td>
</tr>
<tr>
<td>Corn</td>
<td>2.7065644</td>
<td>2.7461595</td>
<td>2.7621547</td>
<td>2.7798875</td>
<td>2.8008799</td>
<td>2.8698510</td>
<td>3.0358627</td>
</tr>
<tr>
<td>Sunflower</td>
<td>0.9352570</td>
<td>0.9359759</td>
<td>0.9360951</td>
<td>0.9361143</td>
<td>0.9361382</td>
<td>0.9670020</td>
<td>1.0097617</td>
</tr>
</tbody>
</table>


