# SENSITIVITY OF 4M MAIZE MODEL TO THE INACCURACY OF WEATHER AND SOIL INPUT DATA

N. FODOR
*e-mail: fodornandor@rissac.hu*

G. J. KOVÁCS

*Research Institute for Soil Science and Agricultural Chemistry of HAS,*
*H-1022 Budapest, Herman O. út 15, Hungary*
*(phone: +36-1-225-3203)*

(Received 14th May 2003; accepted 26th June 2003)

**Abstract**. The accuracy of a crop model is judged mostly by how precise it is in estimating the production. The preciseness of a crop model is determined on one hand by the authenticity of the algorithms describing the processes of the real world, while on the other hand by the quality of its input data. The goal of this study was to test how sensitive crop models are to errors occurring in the measurement of key weather and soil inputs. The key weather and soil related input data for most crop models are air temperature, solar radiation, precipitation, saturated water content, drained upper limit and wilting point. The level of error in air temperature, radiation, and precipitation measurements were reported to be ±0.2 °C, ±2% and ±3% respectively by the Hungarian Meteorological Service. An other study reported 0.008 $cm^3/cm^3$ level of error in determining the drained upper limit. The question we asked was: 'To what extent do the yield and biomass change due to this level of inaccuracies in weather and soil inputs?' The uncertainty caused by the errors of the measured weather elements was found to be 4.7 and 6.9% for the calculated biomass and yield, respectively. The soil parameter errors caused smaller uncertainties in the simulation results. On an average we got 2.3% uncertainty for the biomass and 3.2% uncertainty for the yield. The effect of weather and soil data uncertainties can both strengthen and weaken each other. In certain cases the uncertainty of the simulation could be over 15% due to errors in weather and soil data.
**Keywords:** *crop modelling, error of measurement, uncertainty of simulation*

## Introduction

For a long time models have played a very important role in the scientific research. The primary purpose of crop models is to describe the processes of the very complex atmosphere–soil–plants system using mathematical tools and to simulate them with the help of computers. The ultimate aim of using crop models however is to answer such questions that otherwise could only be answered by carrying out expensive and time-consuming experiments.

It was only in the 1970's that developments in information technology enabled scientists to create the first crop model software using the accumulated scientific knowledge. Since then crop models have been used in numerous educational and scientific projects.

In the year 2001 a new workshop, called 4M, was set up within the Hungarian Soil Science Society with the specific purpose of creating an easy to use package for modelling cropping systems. The 4M package was developed by Hungarian scientists from various institutes in the country. The CERES model was chosen to be a starting point for this project, since it has an open source code and several studies have proved its competence in describing the soil–plant–atmosphere system [3, 4, 5]. 4M is


ISSN 1589 1623

continuously being developed. During the past one and a half years more than ten new subroutines have been incorporated into the model, for instance a new water balance module and a simple weather generator.

The accuracy of a crop model is judged mostly by how precise it is in estimating the production. It is especially significant in cases when crop models are used in decision support. The preciseness of a crop model is determined on one hand by the authenticity of the algorithms describing the processes of the real world, while, on the other, hand by the quality of its input data. Even the 'perfect' model would not be able to simulate the real processes precisely if inaccurate input data were fed into it.

The goal of this study was to test how sensitive crop models are to errors occurring in the measurement of key weather and soil inputs. The key weather and soil related input data for most crop models are air temperature, solar radiation, precipitation, saturated water content, drained upper limit and wilting point. For the test we needed a model, quality weather and soil input datasets, as well as studies on the measurement errors.

## Materials and methods

4M crop model [2] was used for the test. Since the basis of 4M is the well-known CERES model that was used as a starting point for several other crop models in the world, the result of this test gives relevant information for a wide range of scientists. The effect of inaccuracies of the two input data groups were investigated separately, followed by an examination of some error combinations.

### *Inaccuracies of weather inputs*

The weather input was a carefully selected quality dataset of 20 years (1968–87) from a weather centre near the city of Budapest, Hungary. In the past 10 years MILOS 500 and QLC 50 automated equipments were set up at more then 60 different stations of the Hungarian Meteorological Service. A study on the measurement error of these equipments was published recently [6]. The level of measurement errors of air temperature, radiation, and precipitation are presented in *Table 1*.

***Table 1.*** *Level of error of different weather measurements.*

| Data | Level of error |
|---|---|
| Global radiation | ± 2% |
| Temperature | ± 0.2 $^{o}$C |
| Precipitation | ± 3% |

The question we asked was: 'To what extent do the yield and biomass change due to this level of inaccuracy in weather input?'. For setting a reference point, we supposed that the measured dataset ('base dataset' in the following) had no measurement error. Twenty-six additional, 20 year-long weather datasets were created by modifying the components of the base dataset with the corresponding level of errors (*Table 2*).

For example, one weather dataset was created of which daily global radiation data were 2% higher, temperature and precipitation data were 0.2 $^{o}$C and 3% respectively lower than in the base dataset (combination No. 18 in *Table 2*).


ISSN 1589 1623

***Table 2.*** *Characteristics of the modified weather datasets (combinations).*

| Combination No. | Supposed level of errors with which the daily values of the base dataset were modified during model simulations | | |
|---|---|---|---|
| | Global radiation (%) | Temperature ($^{o}C$) | Precipitation (%) |
| 1 (base) | 0 | 0 | 0 |
| 2 | 0 | 0 | +3 |
| 3 | 0 | 0 | -3 |
| 4 | 0 | +0.2 | 0 |
| 5 | 0 | +0.2 | +3 |
| 6 | 0 | +0.2 | -3 |
| 7 | 0 | -0.2 | 0 |
| 8 | 0 | -0.2 | +3 |
| 9 | 0 | -0.2 | -3 |
| 10 | +2 | 0 | 0 |
| 11 | +2 | 0 | +3 |
| 12 | +2 | 0 | -3 |
| 13 | +2 | +0.2 | 0 |
| 14 | +2 | +0.2 | +3 |
| 15 | +2 | +0.2 | -3 |
| 16 | +2 | -0.2 | 0 |
| 17 | +2 | -0.2 | +3 |
| 18 | +2 | -0.2 | -3 |
| 19 | -2 | 0 | 0 |
| 20 | -2 | 0 | +3 |
| 21 | -2 | 0 | -3 |
| 22 | -2 | +0.2 | 0 |
| 23 | -2 | +0.2 | +3 |
| 24 | -2 | +0.2 | -3 |
| 25 | -2 | -0.2 | 0 |
| 26 | -2 | -0.2 | +3 |
| 27 | -2 | -0.2 | -3 |

For the required soil inputs, data of a chernozem profile was used that was selected from the database of RISSAC [7, 8]. The genetic parameters of the Pi3978 cultivar (maize) were used as crop specific inputs.

First, the model was run for the 20 years (1968–87) using the base dataset. The calculated biomass and yield outputs were recorded and the results of the following runs were compared to them. Then the model was run with the modified weather datasets. The calculated yearly biomass and yield values were compared to the ones that were obtained by using the base dataset.

***Inaccuracies of soil inputs***

In a study on in situ measurement of soil hydraulic parameters [1] we have found 0.008 $cm^3/cm^3$ uncertainty of the drained upper limit of the samples coming from the very same profiles and from the very same horizons. In the present study we used this level of error for modifying the saturated water content, drained upper limit and wilting point values of the base soil input file to create input files that are loaded with 'measurement error'. Data of the above mentioned chernozem profile was used as a base input file (*Table 3*).

Twenty-six additional soil input files were created by modifying the base dataset. Each modification was applied for all of the horizons of the profile (*Table 4*).

***Table 3.*** *Characteristics of the chernozem profile.*

| Horizon | Thickness (cm) | Saturated water content ($cm^3/cm^3$) | Drained upper limit ($cm^3/cm^3$) | Wilting point ($cm^3/cm^3$) |
|---|---|---|---|---|
| $A_p$ | 23 | 0.442 | 0.327 | 0.146 |
| A | 20 | 0.454 | 0.314 | 0.135 |
| B | 22 | 0.506 | 0.295 | 0.131 |
| BC | 15 | 0.504 | 0.306 | 0.131 |
| C | 100 | 0.473 | 0.282 | 0.118 |

***Table 4.*** *Characteristics of the modified soil datasets (combinations).*

| Combination No. | Supposed level of errors with which the base dataset were modified for model simulations | | |
|---|---|---|---|
| | Saturated water content ($cm^3/cm^3$) | Drained upper limit ($cm^3/cm^3$) | Wilting point ($cm^3/cm^3$) |
| 1 | -0.008 | -0.008 | -0.008 |
| 2 | -0.008 | -0.008 | 0 |
| 3 | -0.008 | -0.008 | +0.008 |
| 4 | -0.008 | 0 | -0.008 |
| 5 | -0.008 | 0 | 0 |
| 6 | -0.008 | 0 | +0.008 |
| 7 | -0.008 | +0.008 | -0.008 |
| 8 | -0.008 | +0.008 | 0 |
| 9 | -0.008 | +0.008 | +0.008 |
| 10 | 0 | -0.008 | -0.008 |
| 11 | 0 | -0.008 | 0 |
| 12 | 0 | -0.008 | +0.008 |
| 13 | 0 | 0 | -0.008 |
| 14 (base) | 0 | 0 | 0 |
| 15 | 0 | 0 | +0.008 |
| 16 | 0 | +0.008 | -0.008 |
| 17 | 0 | +0.008 | 0 |
| 18 | 0 | +0.008 | +0.008 |
| 19 | +0.008 | -0.008 | -0.008 |
| 20 | +0.008 | -0.008 | 0 |
| 21 | +0.008 | -0.008 | +0.008 |
| 22 | +0.008 | 0 | -0.008 |
| 23 | +0.008 | 0 | 0 |
| 24 | +0.008 | 0 | +0.008 |
| 25 | +0.008 | +0.008 | -0.008 |
| 26 | +0.008 | +0.008 | 0 |
| 27 | +0.008 | +0.008 | +0.008 |

The genetic parameters of the Pi3978 cultivar (maize) were used as crop specific inputs. The weather input was the above mentioned quality dataset of 20 years (1968–87) from Budapest, Hungary. First, the model was run using the base soil dataset and then using the modified soil datasets. The model was run for the 20 year-long period with every soil setting. The calculated yearly biomass and yield values were compared to the ones that were obtained by using the base dataset.


ISSN 1589 1623

Finally we investigated four cases where the worst error combinations were applied both to the weather and the soil data, causing the largest differences ('extreme' combinations) compared to the run that used the base weather and soil dataset. The aim was to find out whether the errors of different types of input data strengthen or weaken each other's effect. We selected two 'extreme' weather combinations (*Fig. 4*). Using combination No. 20 we got consistently lower yields than using the base weather dataset, and using combination No. 18 we got consistently higher yields than using the base weather dataset. Similarly we selected two 'extreme' soil combinations (*Fig. 8*) with which the model gave consistently lower (No. 21) and higher (No. 25) yields than using the base soil dataset. This resulted in four extra combinations. We compared the model results obtained with these extra combinations to those that were obtained by using the base weather and soil datasets.

## Results

### *Effects of inaccuracies of weather inputs*

The differences between the model results obtained by using the base dataset and the modified datasets were calculated for every year and for every combination. Yearly (*Figs. 1–2*) and the overall (for the 20 years) average values (*Figs. 3–4*) of these differences were also calculated.

The simulated yield was more sensitive to errors of the weather measurements than the biomass. If one has no prior knowledge of the weather and of the measurement error, which is usually the case, the expected error/uncertainty of the model results would be 6.0 ± 0.9% ($\alpha = 0.05$) considering the yield, and 3.2 ± 0.3% ($\alpha = 0.05$)

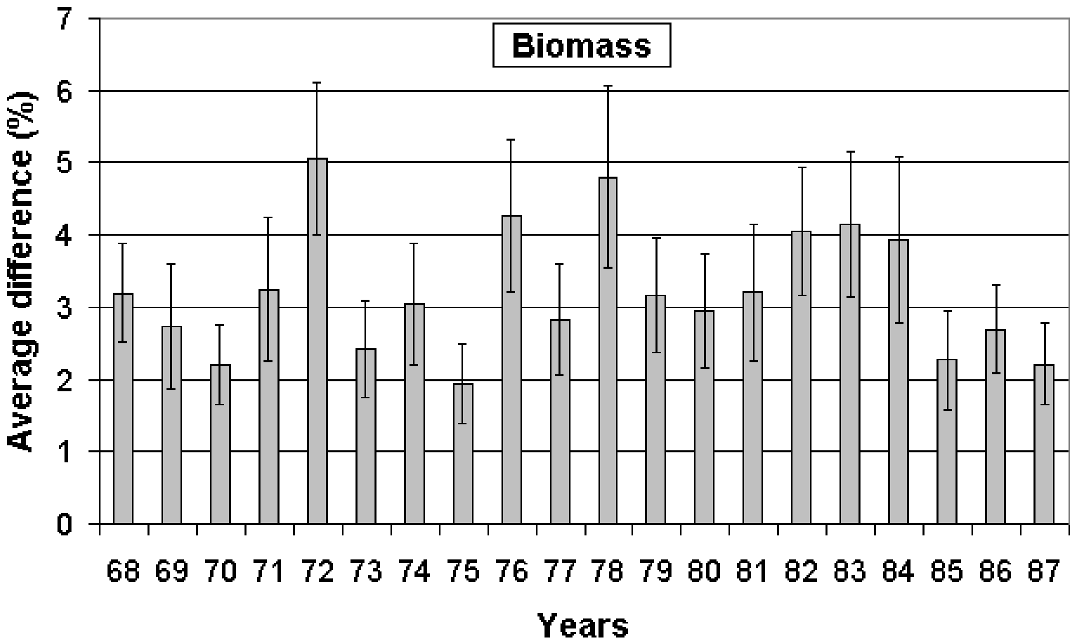


***Figure 1.*** *Yearly averages of the differences between the model results obtained by using the base weather dataset and the 26 modified datasets (combinations). Error bars stand for the 95% confidence interval.*

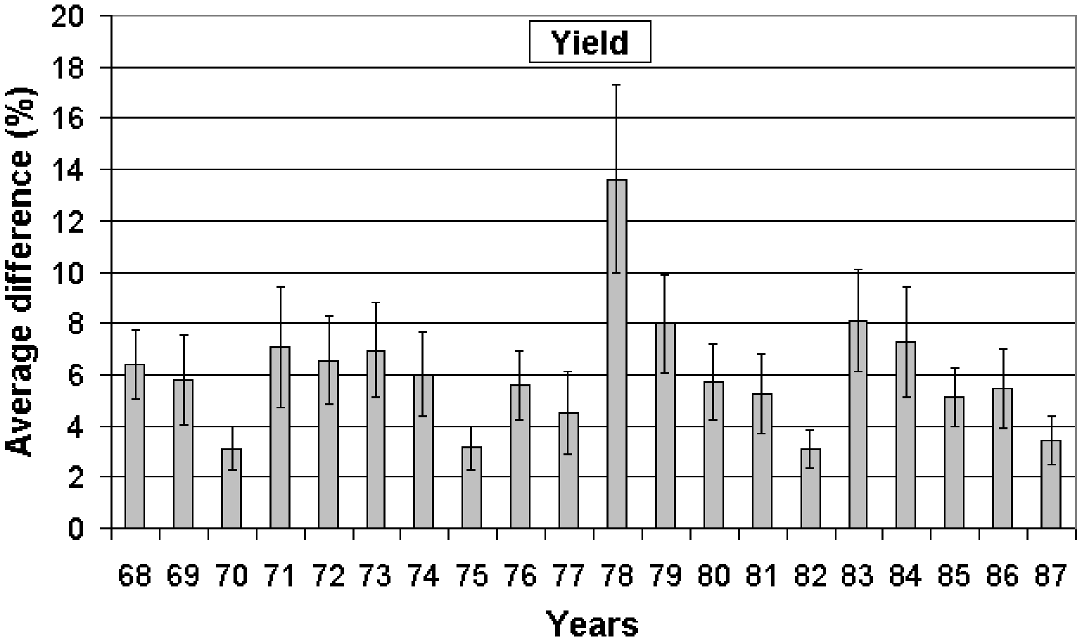


***Figure 2.*** *Yearly averages of the differences between the model results obtained by using the base weather dataset and the 26 modified datasets (combinations). Error bars stand for the 95% confidence interval.*

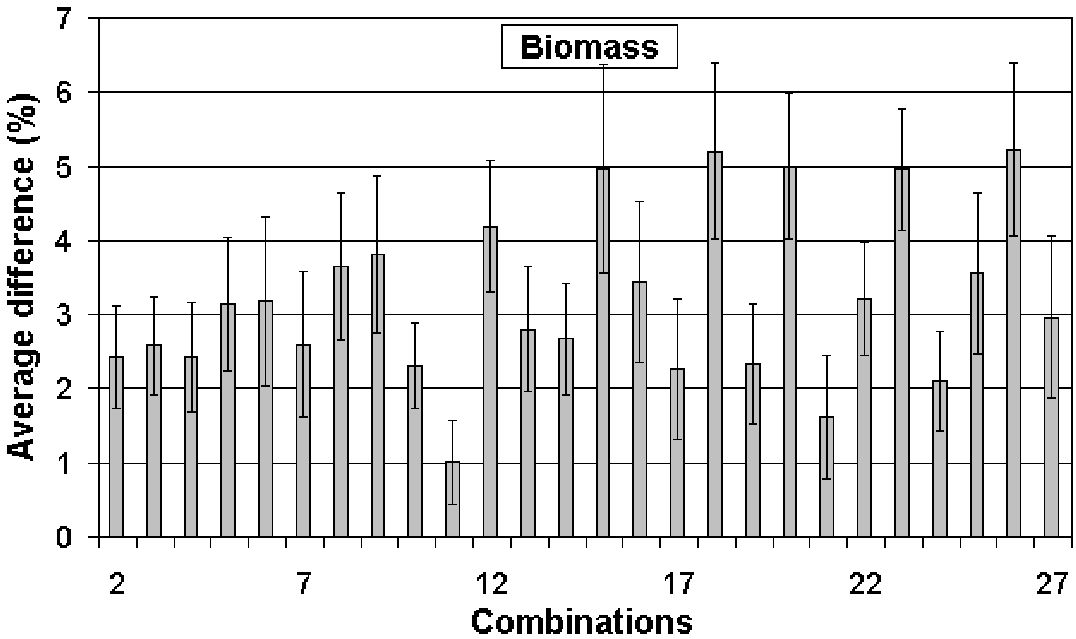


***Figure 3.*** *Average differences (for 20 years) between the model results obtained by using the base weather dataset and the 26 modified datasets (combinations). Error bars stand for the 95% confidence interval.*


ISSN 1589 1623

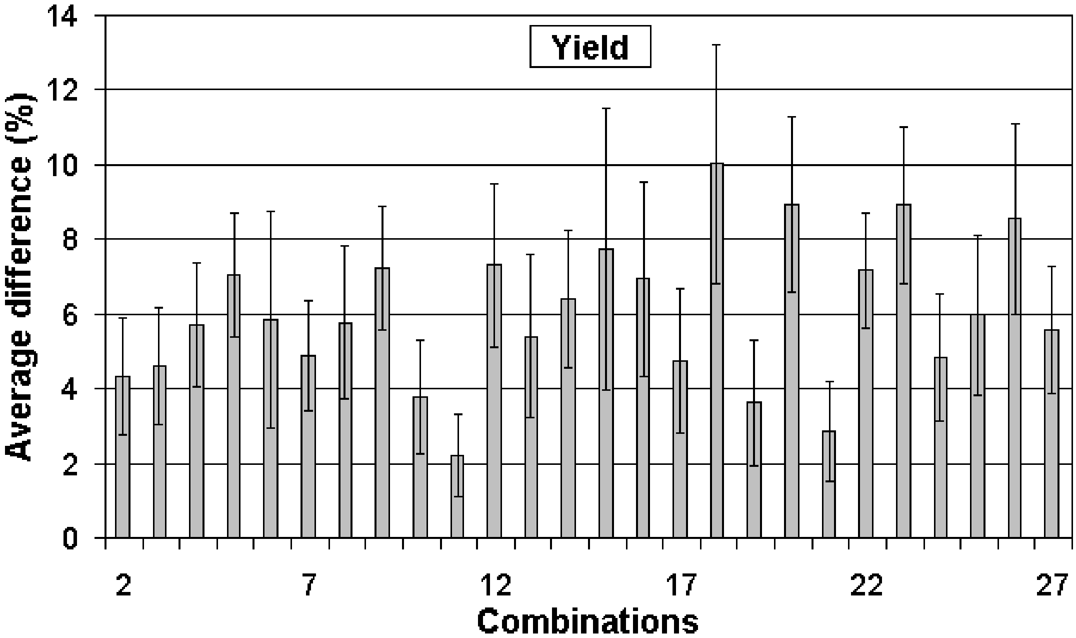


***Figure 4.*** *Average differences (for 20 years) between the model results obtained by using the base weather dataset and the 26 modified datasets (combinations). Error bars stand for the 95% confidence interval.*

considering the biomass. For the simulated yield this uncertainty is 4.7 ± 1.2% ($\alpha$ = 0.05) and 6.9 ± 1.4% ($\alpha$ = 0.05), in high yielding and low yielding years, respectively. For the simulated biomass this uncertainty is 3.1 ± 0.8% ($\alpha$ = 0.05) and 3.3 ± 0.4% ($\alpha$ = 0.05), in high yielding and low yielding years respectively. The inaccuracy of the calculated yield caused by the errors of the measured weather data is significantly larger in low yielding years than in high yielding years.

There might be years when the expected value of error of the calculated yield is around 10% or even higher (1978, 1979 and 1983 in *Fig. 2*). Furthermore some combinations of measurement errors in certain years could cause more than 20% 'shift' in the calculated yield compared to the ones that were obtained by using the 'errorless' base dataset (*Table 5*). Even in 20 year average, there was a combination (No. 18) found to cause more than 10% error in the model calculations (*Fig. 4*). This and three other combinations (No. 20, 23 and 26) were found to be quite 'extreme'. That implies that simultaneous over measuring of the radiation and under measuring of the temperature and of the precipitation (combination No. 18) can cause huge errors in calculating the yield if such weather datasets are used in modelling. If feasible these combinations must be avoided in measuring weather data.

***Table 5.*** *Largest differences between the calculated yield using the base dataset and a modified one (combination).*

| Year | Combination | Difference (%) |
|---|---|---|
| 1971 | 6 | 23.3 |
| 1977 | 20 | 20.2 |
| 1978 | 18 | 36.1 |
| 1979 | 15 | 22.2 |


ISSN 1589 1623

***Table 6.*** *Smallest differences (20 year average) between the calculated yield and biomass using the base dataset and a modified one (combination).*

| Combination | Difference (%) | |
|---|---|---|
| | Biomass | Yield |
| 11 | 1.0 | 2.2 |
| 19 | 2.3 | 3.6 |
| 21 | 1.6 | 2.8 |

On the other hand, there are combinations that were found to cause relatively small errors in the calculations in 20 year average (*Table 6*). This means that the combinations – when the radiation and the precipitation are simultaneously over measured (combination No. 11) or under measured (combination No. 21) – cause only very small errors in calculating the biomass or the yield.

***Effects of inaccuracies of soil inputs***

The differences between the model results obtained by using the base dataset and the modified datasets were calculated for every year and for every combination. Yearly (*Fig. 5–6*) and the overall (for the 20 years) average values (*Fig. 7–8*) of these differences were also calculated.

The simulated yield was more sensitive to errors of the soil parameters than the biomass. The soil parameter errors/uncertainties caused smaller uncertainties in the simulation results than the weather measurement errors. The expected error/uncertainty of the model results would be 3.2 ± 0.6% ($\alpha$ = 0.05) considering the yield, and 2.3 ± 0.4% ($\alpha$ = 0.05) considering the biomass. For the simulated yield this uncertainty is 2.7 ± 1.4% ($\alpha$ = 0.05) and 3.6 ± 0.7% ($\alpha$ = 0.05), in high yielding and low yielding years, respectively. For the simulated biomass this uncertainty is 2.6 ± 0.5% ($\alpha$ = 0.05) and 2.0 ± 0.9% ($\alpha$ = 0.05), in high yielding and low yielding years, respectively.

The five most ‘extreme’ error combinations that caused more than 3% difference in biomass, and more than 5% difference in yield calculations compared to the base dataset had one thing in common. Their drained upper limit values were lower and their wilting point values were higher than in the base dataset (combinations: No. 3, 12 and 21) or their drained upper limit values were higher and their wilting point values were lower than in the base dataset (combinations No. 16 and 25). For these combinations – for the whole profile – the extractable water capacity was 28.8 mm smaller or larger than in the base dataset. This difference resulted in relatively big differences in the simulated yield and biomass. For the rest of the combinations – except for No. 7 – the extractable water capacity was the same as in the base dataset. This result shows the sensitivity of the model to the extractable water capacity.

***Combined effects of inaccuracies of weather and soil inputs***

Effects of weather and soil data inaccuracies and their combined effect are presented in *Table 7*.

The errors of weather and soil measurements can either strengthen or weaken each other’s effect during model simulations. The bad news is that there can be error combinations (weather: 21 combined with soil: 25) where the average difference for 20 years is over 15% compared to the results obtained by using the base weather and soil datasets.


ISSN 1589 1623

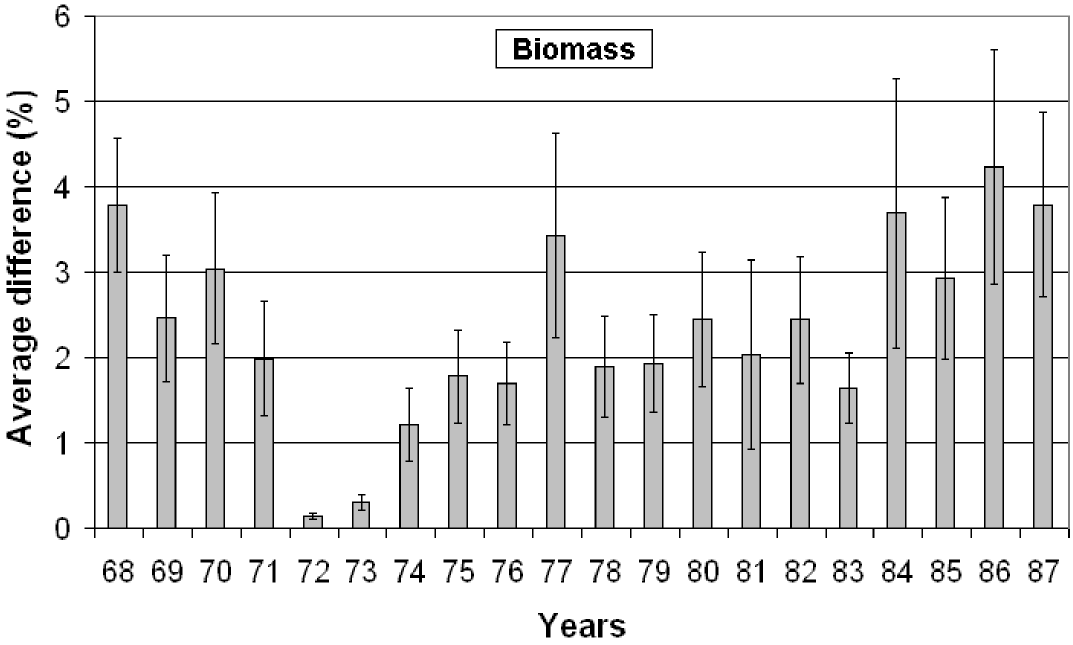


*Figure 5. Yearly averages of the differences between the model results obtained by using the base soil dataset and the 26 modified datasets (combinations). Error bars stand for the 95% confidence interval.*

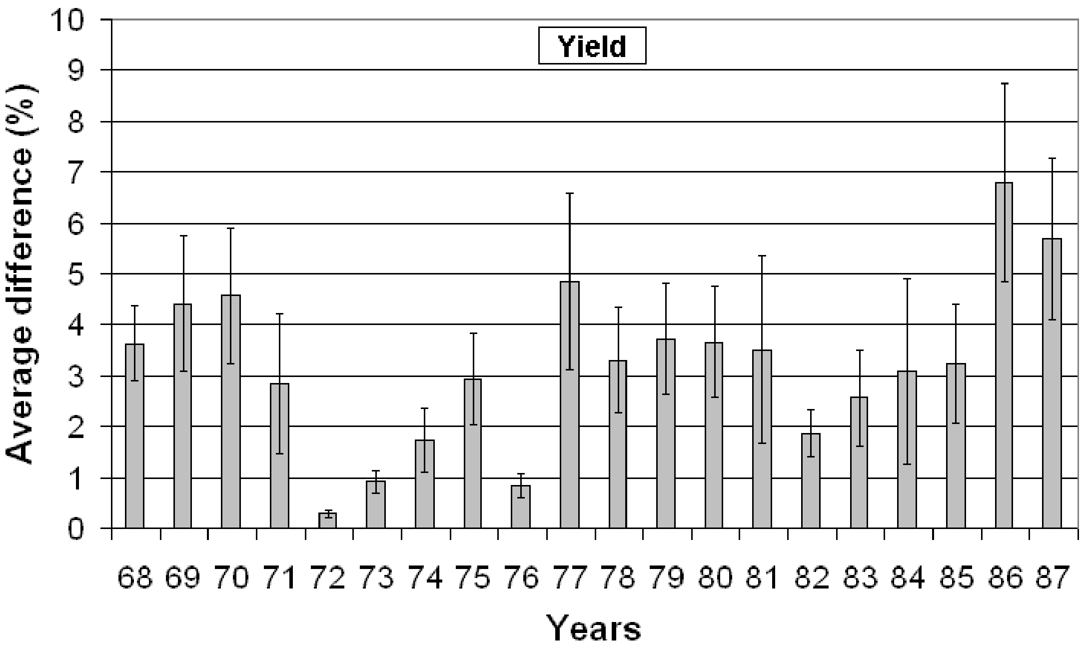


*Figure 6. Yearly averages of the differences between the model results obtained by using the base soil dataset and the 26 modified datasets (combinations). Error bars stand for the 95% confidence interval.*


ISSN 1589 1623

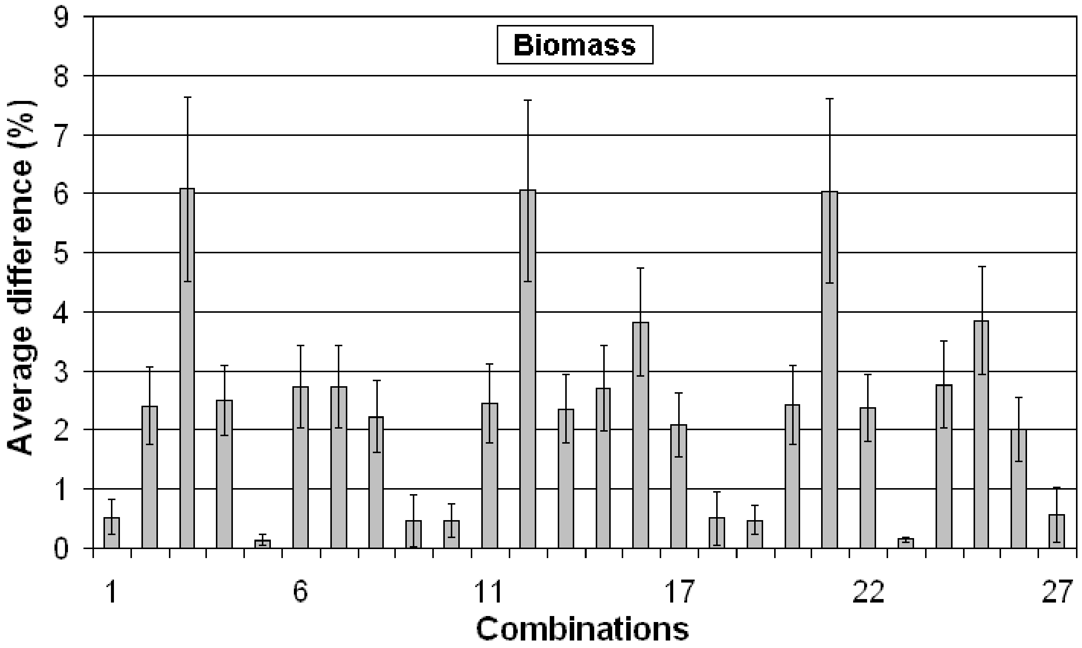


***Figure 7.*** *Average differences (for 20 years) between the model results obtained by using the base soil dataset and the 26 modified datasets (combinations). Error bars stand for the 95% confidence interval.*

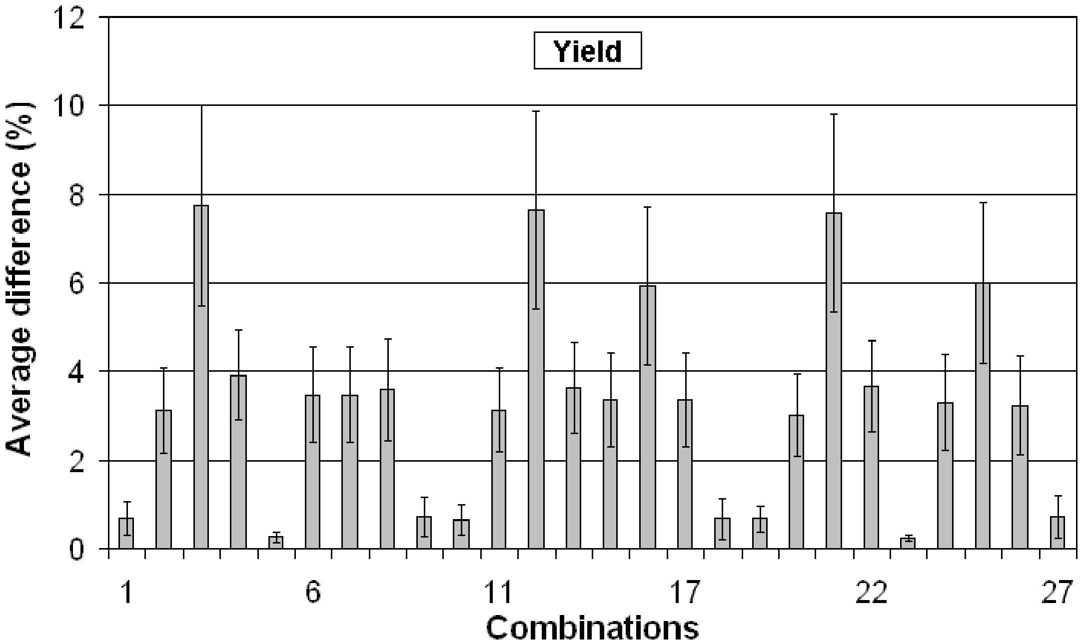


***Figure 8.*** *Average differences (for 20 years) between the model results obtained by using the base soil dataset and the 26 modified datasets (combinations). Error bars stand for the 95% confidence interval.*


ISSN 1589 1623

***Table 7.*** *Effects and combined effects of input data inaccuracies. Average differences (for 20 years) between the simulated yields that were obtained by using the given input data combinations and the yield that was obtained by using the base weather and soil datasets.*

| Combinations | Direct effects ▼ ► | Weather: 21 | Weather: 18 |
|---|---|---|---|
| | | 8.9% | 10% |
| Soil: 21 | 7.6% | 6.4% | 12.8% |
| Soil: 25 | 6% | 15.1% | 7.1% |

Based on this study we can say that it is not only the level of error of the model input data that needs to be given in modelling papers but the uncertainties of the model estimations caused by the errors of the input data, as well. The uncertainty caused by the errors of the measured weather elements was found to be 4.7 and 6.9% for the calculated biomass and yield, respectively. The soil parameter errors/uncertainties caused smaller uncertainties in simulation results. We got 2.3 and 3.2% for the simulated biomass and yield, respectively. Since these two effects can both strengthen and weaken each other we can conclude that the inaccuracies of weather and soil measurements can cause 5–6% uncertainty in the simulated yield on an average. In certain cases the uncertainty of the simulation could be over 15% due to errors in weather and soil data.

**Acknowledgements.** This study was supported partly by grants of OTKA T032768 and T042996.

## REFERENCES

[1] Fodor, N. (2002): A nedvességforgalom modellezése növénytermesztési modellekben. Ph.D. doktori értekezés, Debrecen.
[2] Fodor, N., Máthéné-G., G., Pokovai, K.& Kovács, G.J. (2002): 4M – software package for modelling cropping systems. – European J. of Agr. 18: 389-393.
[3] Jamieson, P.D., Porter, J.R., Goudriaan, J., Ritchie, J.T., van Keulen, H. & Stol, W. (1998): A comparison of the models AFRCWHEAT2, CERES-Wheat, Sirius, SUCROS2, and SWHEAT with measurements from wheat grown under drought. – Field Crop Research 55: 23–44.
[4] Kovács, G.J., Németh, T. & Ritchie, J.T. (1995): Testing simulation models for the assessment of crop production and nitrate leaching in Hungary. – Agricultural Systems 49: 385–397.
[5] Németh, T. (1996): Talajaink szervesanyag-tartalma és nitrogénforgalma. MTA TAKI, Budapest.
[6] Nyitrai, L. (2000): Földfelszíni megfigyelések automatizálása napjainkban. – Légkör 45: 21–26.
[7] Rajkai, K., Várallyay, Gy., Pacsepszkij, J.A. & Cserbakov, R.A. (1981): pF-görbék számítása a talaj mechanikai összetétele és térfogattömege alapján. – Agrokémia és Talajtan 30: 409–438.
[8] Várallyay, Gy. (1987): A talaj vízgazdálkodása. II. Kötet. Akadémiai doktori értekezés, Budapest.


ISSN 1589 1623