DOI:10.28974/idojaras.2022.4.3

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 126, No. 4, October – December, 2022, pp. 481–510

Statistical method for estimating average daily wind speed during the day

Károly Tar[1,*], **István Lázár**[1], and **István Hadnagy**[1,2]

[1]*Department of Meteorology, University of Debrecen, Egyetem Square 1, H-4010 Debrecen, Hungary*

[2]*Department of Biology* and *Chemistry Ferenc Rákóczi II. Transcarpathian Hungarian Institute Kossuth Square 6, UA-90202 Berehove, Ukraine*

*Corresponding author E-mail: tarko47@gmail.com

(Manuscript received in final form January 12, 2022)

Abstract— Meteorologists keep searching and running models to provide the most accurate forecast of wind speed in addition to gaining a more detailed understanding of the wind conditions in Hungary. Wind speed and wind energy estimates, forecasts, and their verification are based on wind statistics from a longer or shorter previous period. Consequently, in addition to dynamic methods, purely statistical models also play an important role, i.e., findings that can be obtained from the statistical analysis of the existing database of measured data. The successive phases of the statistical method for producing scientific or operational information that can be extracted from measured, corrected, and stored meteorological data are generally: statistical analysis/processing, creating, verification, and application of the model, recording of the required information. The targeted information in this paper is the daily average of hourly wind speeds. The exact average of this time series can only be determined after the last measurement. To estimate this average during the day, however, the so-called sliding average model has been developed, which can be applied to any climatic element if its measured values are recorded at regular times over a certain period of time. The results presented in this paper are recommended for the preparation of the so-called "timetable", which is one of the most difficult problems for wind farm operators. This is basically the estimation of the amount of electricity produced the following day over short periods. It would be a significant help in the above if we can determine the probability of a decrease or increase in the average wind speed on the next day (and with it, the average daily wind power), or which of these two probabilities is greater. This requires an estimate of average wind speed of the next day. In addition, the results of one of our previous studies on the statistical structure of day-to-day changes in average daily wind speeds were also used. According to the results of the monthly testing of the model over a given period, the frequency of good estimates is between 80.6 % and 54.8%.

Key-words: sliding average model, wind statistics, wind farms, daily wind power, event frequency, Hungary

1. Introduction, antecedents

Wind speed and wind energy estimates, forecasts, and their verification are based on wind statistics from a longer or shorter previous period. Consequently, in addition to dynamic methods, purely statistical models also play an essential role. An overview of these can be found in the work of *Aggarwal* and *Gupta* (2013).

A statistical method for producing scientific or operational information that can be extracted from measured, corrected, and stored meteorological data is presented here. The successive phases of this are: statistical analysis/processing, modeling, model verification, model application, recording of the desired information.

The targeted information here is the average of the values of a climatic element measured at regular times. The exact average of this time series can only be determined after the last measurement. However, in some cases, it may be necessary to estimate this value with an acceptable error before the last measurement.

When wind energy is harnessed, this particular climatic element is wind speed. With the integration of wind energy into electricity grids, it is becoming increasingly important to obtain accurate wind speed/power forecasts. Accurate wind speed forecasts are necessary to schedule dispatchable generation and tariffs in the day-ahead electricity market *(Bremnes et al.,* 2002; *Kavasseri* and *Seetharaman*, 2009; *Shukur* and *Lee*, 2015). A very important element of this process is the preparation of a so-called "timetable", a difficult problem for wind power plant operators. This is basically the estimation of the amount of electricity produced the following day over short periods. It would be a significant help in the above if we can determine the probability of a decrease or increase in the average wind speed on the next day (and with it, the average daily wind power), or which of these two probabilities is greater. This requires an estimate of average wind speed of the next day.

In two previous studies (*Tar* and *Lázár*, 2018, *Tar*, 2021), we described the process of building a mathematical statistical model that is ultimately suitable for estimating the sign of the next day's average wind speed change and the magnitude of the average wind speed of the next day from today's average wind speed. The most important steps in the construction of this model are briefly summarized in the following.

The model is based on a time series of observed average daily wind speeds transformed to the height of 10 metres. Analyses were performed on the entire time series and its subsets of days for cyclone and anticyclone macrosynoptic situation groups (*Péczely*, 1961) and their transitions.

The changes of daily average wind speed from day to day were characterized by the relative value of

$$\Delta v_r = \frac{v_n - v_p}{v_p}, \qquad (1)$$

where v_p is the average wind speed of the present day and v_n is that of the following day. The value of Δv_r is more or less independent of the height of the measurement, i.e., the height of the anemometer, error is made only for the day preceding the change of the measurement height. When given in percentage, it shows the change of the daily average wind speed of the next day in relation to that of the previous day.

Since Δv_r is the observed value of a random variable with a special structure, its most important statistical functions have been analyzed in more detail, primarily in relation to the situational groups and their transitions, using the average daily wind speeds of nine Hungarian meteorological stations over 10 years (1991–2000).

We looked at the relationship of the sign of the relative change and the average wind speed of the present day. According to Eq. (1), Δv_r has a very complex function relationship with the average wind speed (v_p) of the present day, given that the average wind speed (v_n) of the next day also depends on it. Therefore, it is advisable to consider the (v_p,Δv_r) relationship as stochastic. Logarithmic regression was the closest correlation. The correlation index i(v_p,Δv_r) showing the closeness of logarithmic regression varies between 0.404 and 0.592, with its highest valuemeasured at the meteorological station Győr at the transition from the day in the cyclone situation group to the same situation group (CG/CG). *Fig. 1* shows the regression calculated for Debrecen station for the entire 10-year period.

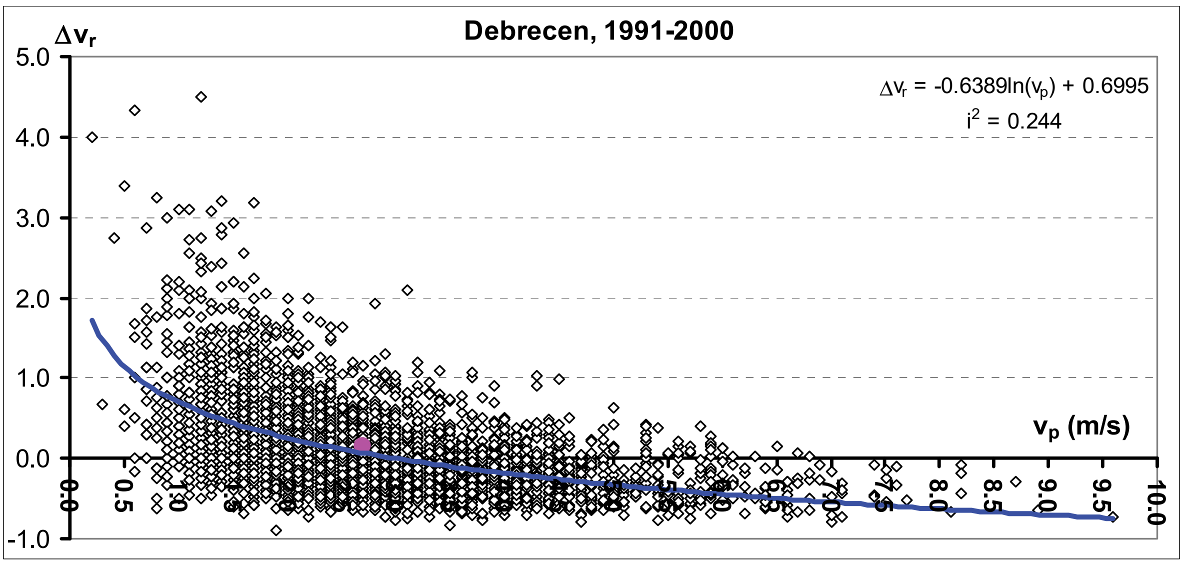

Fig. 1. Logarithmic regression between today's average wind speed (v_p) and next day's relative change (Δv_r) in Debrecen.

In all cases, the regression curve intersects the horizontal axis, i.e., the v_p axis. Let this be the zero point, v_{p0}. For the (x,y) coordinate points of the regression curve, $y>0$ before the zero point and $y<0$ after the zero point. Therefore, it can be assumed that the sign of the observed values of Δv_r may also be associated with zero points. Zero points can therefore be considered as threshold values for the examination of the sign of Δv_r. Detailed analysis confirmed this, but calculating v_{p0}, as we have seen, is not simple, consequently it is advisable to use a statistic that is easier to calculate or may be known already as a threshold instead. Citing previous studies, we presumed that average wind speeds in the categories ($[v]$) could be used as thresholds instead of zero points.

Based on the results of the detailed analysis related to this, it can be stated that if the average wind speed of the present day is less than the average speed of the category, the increase in the average wind speed of the following day is 1.4–2.3 times more likely, on average 1.9 times, than its decrease. If, on the other hand, the daily average wind speed is higher than the average speed in the category, the probability of a decrease in the next daily average wind speed is 1.6 to 5.2 times, on average 2.4 times, greater than that of an increase. Therefore, only $[v]$ depends on the weather situation.

Therefore, in order to make our model usable operationally for estimating the sign of the change in the average wind speed by the next day, the following conditions must be met:

- The average of the long-term wind speed of the site at an altitude of 10 m for the whole period and its selected subsets has to be known.
- The exact average wind speed of the present day, transformed to 10 m height has to be known.

However, the exact average daily speed can only be determined from hourly data at the end of the day. In order to use the estimate, this data shall be known sooner, therefore, an approximate value that can be calculated earlier has to be applied. The method intended to determine this value is presented in the following.

2. The sliding average model

The problem can be generally stated as follows: The measured values of a climatic element are recorded at regular times (e.g., hourly, daily) during periods $i=1, 2, ..., (n-1), n$. The exact average of this time series can only be determined after measurement n. However, in some cases, it may be necessary to estimate this average before date n with an acceptable error.

2.1. Structure of the model

The statistical model to be presented in the following was designed to solve the above problem. The bases of the model were published by: *Tar*, 1990, 1993, 1995ab, 2004, 2019, *Tar* and *Kircsi*, 2001, *Tar* et al., 2001, 2007, *Tar* and *Szegedi*, 2011.

The database of the model is composed of a statistically sufficient measurement data matrix for a given climate element, the elements of which shall be $x_{i,j}$. The general form of the matrix is: j: row index, j = 1, 2, ..., (N-1), N, i: column index, i = 1, 2, ..., (n–1), n. Thus, N can represent the number of days involved in processing, and n can be the number of measurements at equal intervals (e.g., hourly) per day.

At each measurement time i, the $[x_{i,j}]$ elements of the so-called *sliding averages* matrix are counted per row j:

$$[x_{i,j}] = \frac{1}{i}\sum_{k=1}^{i} x_{k,j} \ . \tag{2}$$

Thus, $[x_{i,j}]$ represents the average calculated up to the measurement time i of row j, i.e., $[x_{n,j}]$ gives the total average of the row j. Knowing this, the so-called *relative sliding averages* are obtained as

$$R_{i,j} = \frac{[x_{i,j}]}{[x_{n,j}]} , \tag{3}$$

which shows that the average until time i is the proportion of the average of the complete row. Their average – the so-called *average relative sliding average* – has to be calculated at each measurement time:

$$[R_i] = \frac{1}{N}\sum_{j=1}^{N} R_{i,j} \ . \tag{4}$$

An example of deriving the above parameters is given in *Table 1*, where $x_{i,j}$ is the hourly (i) wind speed on day j of the data matrix measured at Debrecen meteorological station on July 20, 1991. The process on one day is presented in *Fig. 2*.

Table 1. Relative sliding average parameters of the wind speed data matrix measured at Debrecen station on July 20, 1991

i	$x_{i,j}$	$[x_{i,j}]$	$R_{i,j}$
1	2.9	2.90	0.78
2	2.9	2.90	0.78
3	1.8	2.53	0.68
4	2.9	2.63	0.71
5	3.4	2.78	0.75
6	3.3	2.87	0.77
7	4.2	3.06	0.83
8	4.7	3.26	1.88
9	4.9	3.44	0.93
10	4.6	3.56	0.96
11	5.4	3.73	1.01
12	5.6	3.88	1.05
13	6.1	4.05	1.10
14	6.1	4.20	1.14
15	5.8	4.31	1.16
16	4.7	4.33	1.17
17	4.2	4.32	1.17
18	2.8	4.24	1.15
19	1.4	4.09	1.11
20	0.8	3.93	1.06
21	0.8	3.78	1.02
22	2.8	3.73	1.01
23	2.9	3.70	1.00
24	3.5	3.69	1.00

$R_{i,j}$ and thus $[R_i]$ do not depend on the height of the device, because they are relative quantities. Their value is also not disturbed by changes in the height of the device, if it can be taken as constant within a series (e.g., one day). However, $[R_i]$ depends on the selected climate element and is presumably dependent on the location of the observation, the weather situation, as well as the season. Therefore, it is advisable to produce the average relative sliding average at a given location in addition to the entire database for certain subsets of this, e.g., by macrosynoptic position group or situation, for the growing season, seasonally, etc.

The $[R_i]$parameter is used for testing the model and of course for its operative running, i.e., for the estimation of the exact series average outside database of the given climatic element used in the formation of the model. The row averages $[x_{n,j}]$ (e.g., the daily averages) are estimated from the sliding average $[x_{i,j}]$ at measurement time i. The estimation is made by using the average of $R_{i,j}$ instead of $R_{i,j}$ in Eq.(3), i.e.,

$$[x_{n,j}]_{estim,i} = \frac{[x_{i,j}]}{[R_i]} \quad . \tag{5}$$

Therefore, the estimated value of the total average also depends on the time point from which the estimate is made.

Of course, a different parameter of R_i distribution selected for the particular goal (i.e., mode) can also be used instead of $[R_i]$ in the course of the estimation.

In the course of the *verification*, Eq.(5) is performed at all times of all series, then – as $[x_{n,j}]$ is known – the $[E_i]$ average of the relative error of the estimations (in %) at the times is calculated from the relative error per estimate, $E_{i,j}$

$$E_{i,j} = 100 \frac{|[x_{n,j}]_{estim,i} - [x_{n,j}]|}{[x_{n,j}]} \quad , \tag{6}$$

hence

$$[E_i] = \frac{1}{N} \sum_{j=1}^{N} E_{i,j}. \tag{7}$$

Eq.(6) measures the magnitude of the daily relative error, which is always positive or zero. If we also want to examine the sign of the relative error, we use the form of Eq.(6) without absolute value.

The deviation of the approximated or modeled values from the actual values is most often measured by the RMSD (root mean square deviation) parameter (*Armstrong* and *Collopy*, 1992; *Olaofe* and *Folly*, 2012). This number is actually the so-called residual standard deviation, that is, the square root of the mean of the square errors. RMSD is sensitive to outliers, which means that larger errors disproportionately affect its value. On the other hand, RMSD is a measure of accuracy, to compare forecasting errors of different models for a particular dataset and not between datasets, as it is scale-dependent (*Armstrong* and *Collopy*, 1992). Because of these, the relative errors defined by Eq.(6) were chosen to verify the model. Since the estimation given by Eq.(5) is performed at each measurement time point (i), the magnitude of the first few errors would disproportionately increase the RMSD value. For this reason, the trend of changes in errors over time (necessary reduction) would not be clear either. On the other hand, due to the scale dependence of RMSD, it is not possible to compare the usability of the model for different climatic elements. However, the use of relative values reduces the dependence of the error rate on the size and number of sample elements. This makes the $[E_i]$ parameter more comparable and increases the information content of the conclusions that can be drawn.

For actual series-by-series estimates, errors (Eq.(7)) are determined after calculating the last (n) sliding average, $x_{n,j}$ giving the actual average so that $[x_{n,j}]_{estim,i}$ estimates are stored.

If estimating the sum of the data matrix per series is the aim (e.g., monthly precipitation or global radiation), we use sliding sums instead of sliding averages.

2.2. *Database of model development*

The database of model development is now composed of the hourly wind speeds of five Hungarian meteorological stations: Szombathely, Budapest-Pestszentlőrinc, Debrecen, Szeged (non-mountain stations), and Kékestető (height a.s.l. is 1011.3 m) in the period 1991–2000. In the case of the diagrams showing daily runs, we use the 24-hour schedule accepted in Hungary to mark the times.
The statistical parameters of the model have been determined and verified for the entire period above and for its following subsets: the anticyclone and cyclone situational group of Péczely's macrosynoptic situations (*Péczely*, 1961; *Károssy* 1993, 1998, 2001) and seasonally.

The total number of days in the period 1991–2000 is 3653. Measurements of 78 days are missing in Szombathely and those of 5 days are missing in Budapest. This, however, causes no significant difference between the proportions of neither the situation groups nor the number of days of the seasons compared to the other three stations. The proportion of days in the anticyclone situational group varies between 67.1 and 67.4%, and the proportion of days in each season varies between 24.4% and 25.5%.

2.3. *Specifics of the daily changes of the average relative sliding average*

The hourly sliding averages shall therefore be calculated first from the hourly wind speeds for each day of the whole period or of the above subsets, based on Eq.(2). The 24th hour sliding average is the average wind speed per day *j*; dividing the sliding averages by this, the hourly value of the Eq.(3) relative sliding averages is obtained. The process on one day is shown in *Fig. 2*. This is followed by averaging the relative sliding averages by the hour, i.e., the determination of $[R_i]$ values.

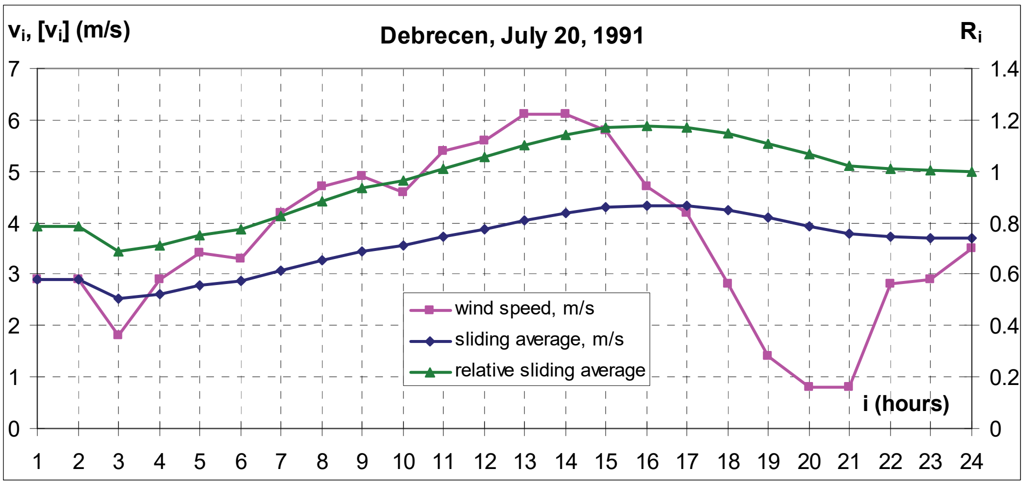

Fig. 2. The values of hourly wind speed (v_i), sliding average ($[v_i]$) and relative sliding average (R_i) in Debrecen on July 20, 1991.

Fig. 3 shows the daily changes of $[R_i]$ at the five stations for the entire period, as well as for the anticyclone and cyclone situation groups and seasons. Hourly differences exceed 0.05 only in the anticyclone situation group in the early hours after midnight. Between about 1pm and 8pm, the decreasing order of Szeged, Debrecen, Szombathely, Budapest is formed in all three cases. This is also more or less observed in the seasons, most notably in autumn.

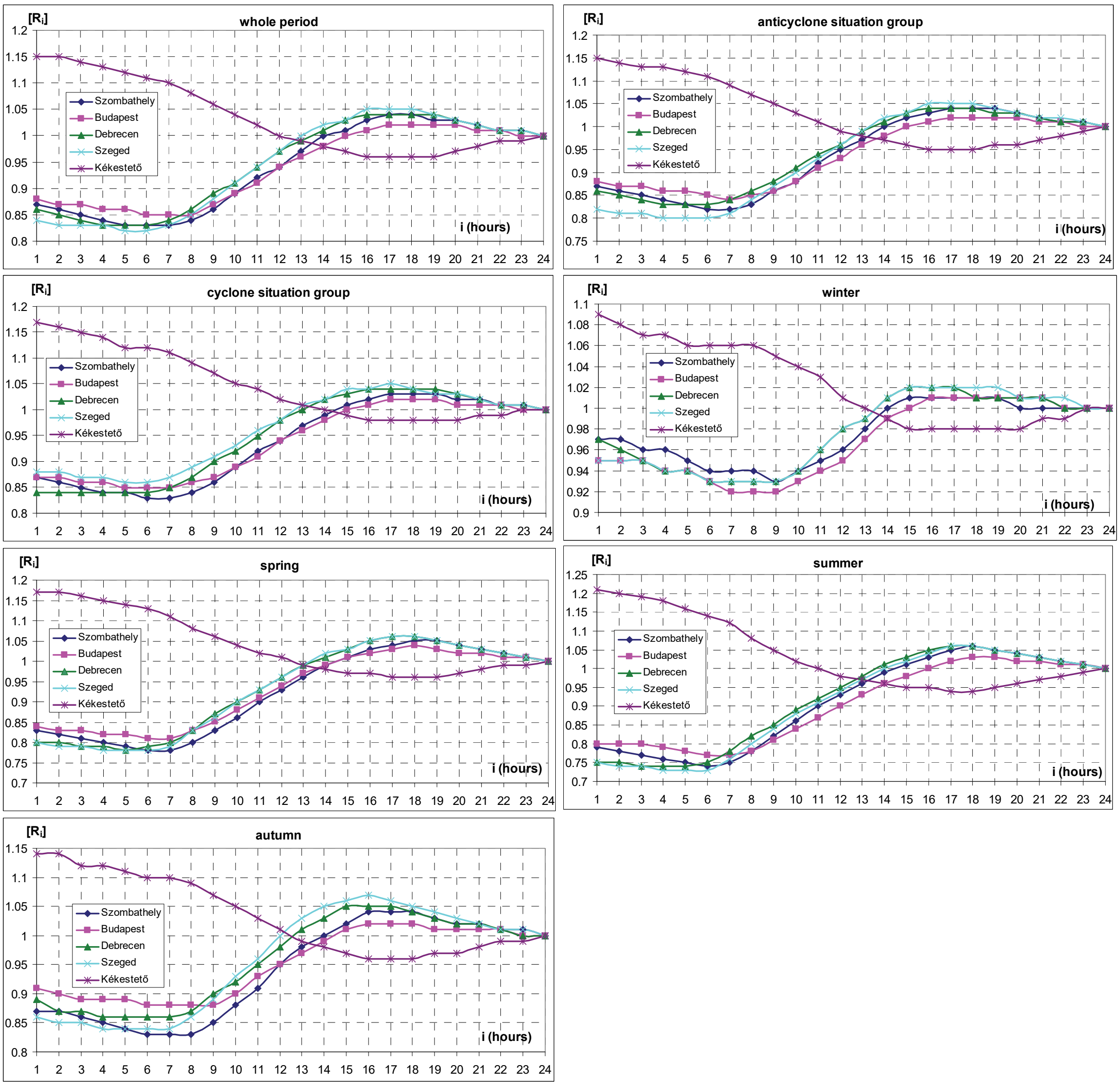

Fig. 3. Daily course of mean relative sliding averages ($[R_i]$).

At the four non-mountain stations, the waves of $[R_i]$ take their lowest values between 4am and 7am ranging from 0.82 to 0.86. The maximums occur at 5pm and 6pm, with values between 1.02 and 1.05. In all three cases, the amplitude (maximum-minimum) of the $[R_i]$ waves is the smallest in Budapest (0.17, 0.18). The highest amplitude was observed in Szeged (0.23, 0.25) during the entire period and in the anticyclone situation group, however, in the cyclone situation group, in addition to the Budapest minimum, amplitudes can be considered equal.

In winter, extreme values persist for a long time. The maximum value (1.01, 1.02) can be observed for 4–6 hours between 3pm and 9pm, and the minimum value (0.92, 0.93) can be observed for 3–4 hours between 6am and 9am. Szombathely is an exception to the latter, with the minimum value occurring at 9am. Amplitudes are practically equal (0.08, 0.09). In spring, the maximum value (1.04-1.06) occurs between 5pm and 7pm and can be observed at all four stations at 6pm. The minimum value (0.78-0.80) occurs between 4am and 7am, and can be observed at 5am or 6am at all four stations. The amplitude is the smallest in Budapest with 0.23, and can be considered equal at the other three stations (0.27, 0.28). In summer, maximums (1.03–1.06) occur everywhere at 6pm, while the minimums (0.73-0.74) occur between 5am and 6am at the four stations.

The amplitude is again the smallest in Budapest, with 0.26, and can be considered equal again at the other three stations (0.32, 0.33). In autumn, the maximum value (1.02–1.07) occurs at 4pm everywhere except Szeged, where it occurs at 5pm and 6pm as well. The minimums (0.83–0.88) occur at 6pm or 7pm at all four stations. Amplitudes are now more diverse: 0.14 in Budapest, 0.19 in Debrecen, 0.21 in Szombathely, and 0.23 in Szeged.

At the altitude of Kékestető, the daily wind speed shows its minimum at early afternoon, consequently, the daily changes of $[R_i]$ are the opposite of that of the other four stations, which are located at much lower altitudes. The maximums occur at 1am or 2am (1.21–1.09) for each of the seven cases. The time of the minimums is spread between 3pm and 8pm, but occurs at 5pm and 6pm in all seven cases. The amplitude is the largest (0.27) in summer and the smallest (0.11) in winter.

2.4. Verification

Using the known time series of sliding averages and average relative sliding averages produced from hourly wind speeds of the present day, the average daily wind speed can be estimated at any hour of the day based on Eq.(5). According to the above, the estimation can be made based on the time series of $[R_i]$ for the whole period and for the macrosynoptic situational group, and also on its seasonal time series.

For the verification of the model, the wind speed database involved in the modeling was used. Since the daily average wind speed is now known, the error of the estimate can be calculated per hour and then the average of these can be calculated as well. The process is illustrated in *Fig. 4*. Average hourly estimate errors calculated on the basis of Eq.(7) are shown in *Fig. 5*.

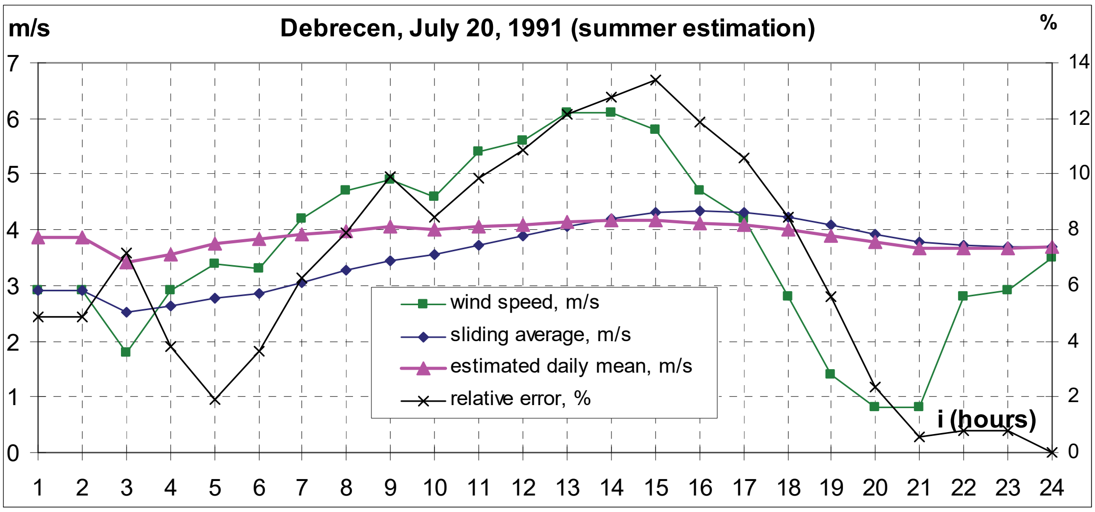

Fig. 4. The estimation process to verify the model.

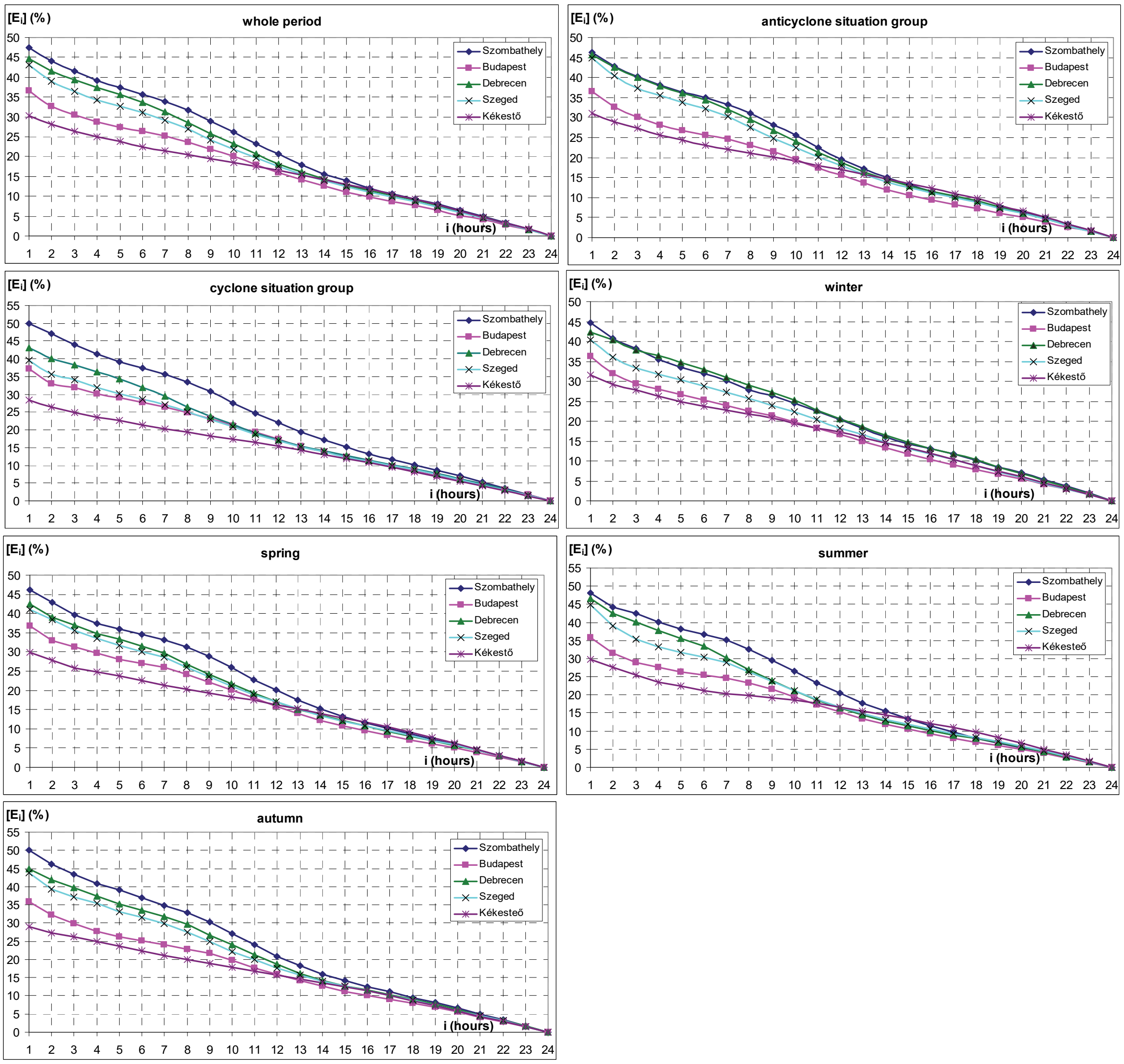

Fig. 5. The daily course of the hourly average relative error of the estimate.

According to *Fig. 5*, the average relative error ($[E_i]$) decreases rapidly as the time of the estimate approaches the end of the day. The values of the $[E_i]$ time series can be approached well with the linear trend. The steepness of the linear trend specifies the value of the average daily decrease, and the absolute value defines the measure. The average hourly decrease has the highest absolute value at Szombathely and the lowest at Kékestető, and the order between them is Debrecen, Szeged, Budapest in all categories. A certain orographic effect can therefore be assumed, since the last three stations are located in a lowland environment, and Kékestető has an even more open horizon. The maximum values at Szombathely are between 2.1 and 2.2%/hour except in spring, and the maximum in spring is 1.85%/hour. The minimum values at Kékestető are between 1.2 and 1.3%/hour. In the other three stations, the average daily decreases are between 2.0 and 1.5 %/h in any category. As a result of the rapid decrease, the values of $[E_i]$ fall below 20% in all cases after 1pm.

Fig. 5 also indicates that if the estimation of the daily averages during the early afternoon is enough, it will not be necessary to subdivide the studied period.

3. Testing the model

For the operational application of the model, the user must have a time series of average relative sliding averages ($[R_i]$) produced from long-standing hourly wind speeds at that location or at a nearby weather station for at least the entire (annual) period. (Under the entire period, we mean at least one year, i.e., in this case we get the annual averages of $[R_i]$.) On a given day, the estimation of the average wind speed of the next day has to be performed based on one or more sliding averages of the period after 12 o'clock. The average of these can also be considered as a good estimate.

It is assumed that the estimated value of the daily average wind speed will approach the true value with the smallest error if the estimate is performed from time *i* where $[R_i] \approx 1$. These times before 4pm – at least 8 hours before the end of the day when it is still worth to perform the estimate – allowing an absolute deviation of 0.01 ($0.99 \leq [R_i] \leq 1.01$) are shown in *Table 2*. The most frequent times listed in the table are 1pm, 2pm, and 3pm.

Table 2. The times (hours) in which $[R_i] \approx 1$

Period	Szombathely	Budapest	Debrecen	Szeged	Kékestető
whole	14, 15	15, 16	13, 14	13	12, 13
anti-cyclonal	14	15, 16	13, 14	13	11, 12
cyclonal	14, 15	15, 16	13	13	13, 14, 15
winter	14, 15, 16	14, 15, 16	13, 14	13, 14	12, 13, 14
spring	14, 15	14, 15	13, 14	13	12, 13
summer	14, 15	16	14	14	11
autumn	14	15, 16	13	12	12, 13

In the following, model simulations using annual SODAR data measured at 30 m height, in Debrecen, in 2013, are discussed. The input – similarly to wind speed measured on wind power plants – is a sequential file, the records of which include the date (year, month, day) and time of measurement (0, 10, 20, 30, 40, and 50 minutes of every hour). The estimate is made at 1pm, 2pm, and 3pm using the average relative sliding averages ($[R_i]$) determined for the whole period (year). The daily pattern of $[R_i]$ values in Debrecen is presented in *Fig. 6*. It can be seen that the values are somewhat larger (by 0.02–0.03) at the three times in autumn than in other cases, where they are nearly equal. This, however, does not mean that estimating the daily change of the average daily wind speed would be independent of the season, because the other parameter of the estimate is the average wind speed for the period.

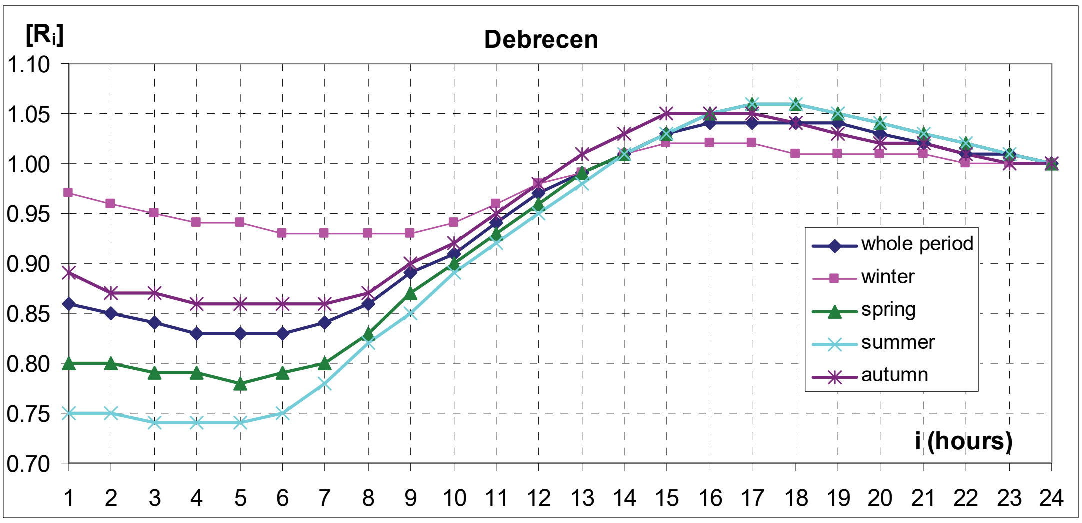

Fig. 6. Annual and seasonal daily course of hourly average relative sliding averages in Debrecen based on SODAR data measured at 30 m height in 2013.

Monthly analyses are also carried out throughout the whole period (year). From these we can infer the reality of the annual pattern of each characteristics (the significance of the differences between them).

3.1. Frequency of accurate, under-, and overestimates

First, we analyze the frequency of the sign of the estimation error, i.e., the frequency of accurate, under-, and overestimation is analyzed. The estimate is considered accurate if its difference from the real daily average is 0.0 to one decimal point. Underestimation and overestimation mean that the difference is negative or positive, respectively. The annual pattern of the proportion of accurate estimates and the differences of underestimates and overestimates are shown in *Fig. 7*.

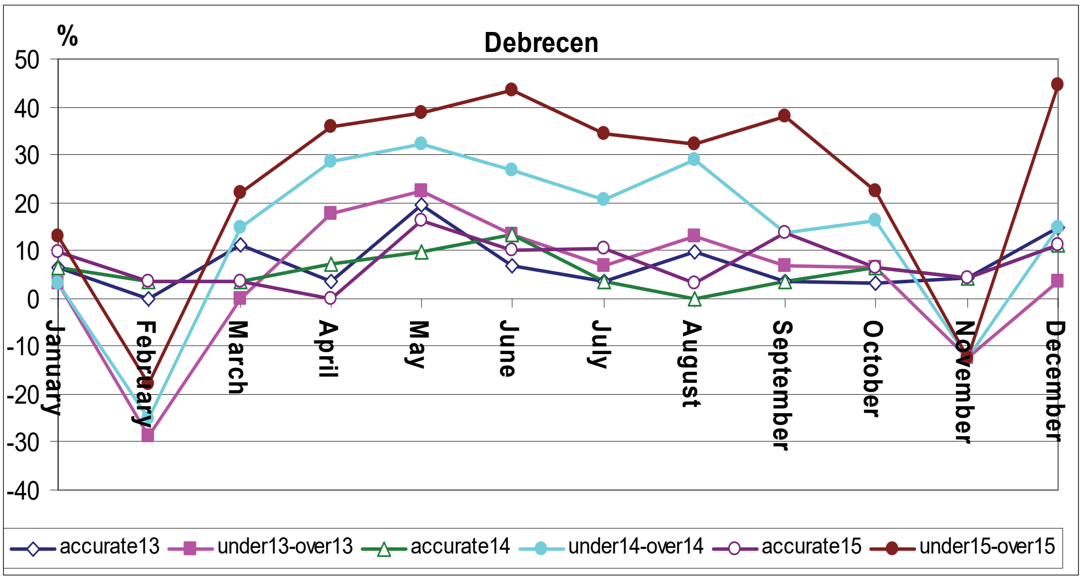

Fig. 7. The annual course of the proportion of accurate estimates and the difference between under- and overestimates. (Note, that 13 refers to 1pm, 14 refers to 2pm, and 15 refers to 3pm.)

Surprisingly, the maximum of the monthly relative frequency of accurate estimates is divided between 1pm and 3pm in almost 50%–50% of the cases. The difference between the monthly ratio of underestimates and overestimates is negative in February and November (i.e., the number of overestimates is higher in these months), but with the exception of November, this difference is the largest for the 3pm estimate. Therefore, the number of underestimates increased as the estimate time increases in the present case. The evaluation of the full-year results of the estimates also shows this (see *Table 3*).

Table 3. Accurate, underestimation and overestimation rates throughout the year

% at	accurate	under-	over-
		estimation	
1 pm	7.2	48.8	43.9
2 pm	6.1	54.0	39.9
3 pm	7.8	58.7	33.5

3.2. Statistics of simple estimate error

The simple measure of the signed estimation error is the difference between the estimated and the actual daily average wind speed, i.e., $[x_{n,j}]_{estim,i} - [x_{n,j}]$, and now i=13, 14, 15, j=1, 2, …., N.

The main statistical characteristics of this error are briefly analyzed for the three estimates for the whole year, paying attention to the fact that the real difference is the absolute value of the error. The values of the most important characteristics are given in *Table 4*.

Table 4. The most important statistical characteristics of daily simple estimation errors (m/s, Debrecen, 2013)

m/s, at	1pm	2pm	3pm
mean	0.00	-0.07	-0.14
st. deviation	0.71	0.64	0.56
maximum	2.80	2.30	1.80
minimum	-2.20	-2.00	-1.80
range	5.00	4.30	3.60
skewness	0.32	0.25	0.16
kurtosis	0.99	0.86	0.76
mode	-0.20	-0.30	-0.20
median	0.00	-0.10	-0.20

The characteristics of the simple estimation error decrease over time, except for the minimum value and mode, and the minimum increases. Modes can also be considered equal. This can be decided based on the frequency distribution of errors, which is prepared by classifying the errors into 0.2 m/s long intervals taking the extreme values into account as well.

The columns of *Fig. 8* show that errors are classified into the interval of (-0.4:0.0) m/s with highest frequencies, 27.7, 31.2, and 32.9% similarly to at all three times, averages and medians. In addition, the values of the skewness and kurtosis coefficients suggest the possibility of an approach with normal distribution. To determine the goodness of the fit, χ^2 test was used. Accordingly, the hypothesis that the frequency distribution of the magnitude of the estimation errors is normal, is not rejected at a significance level of 0.05 for any of the hourly estimates. This means that differences between -0.4 and 0.0 m/s are most likely to occur at all three estimation times.

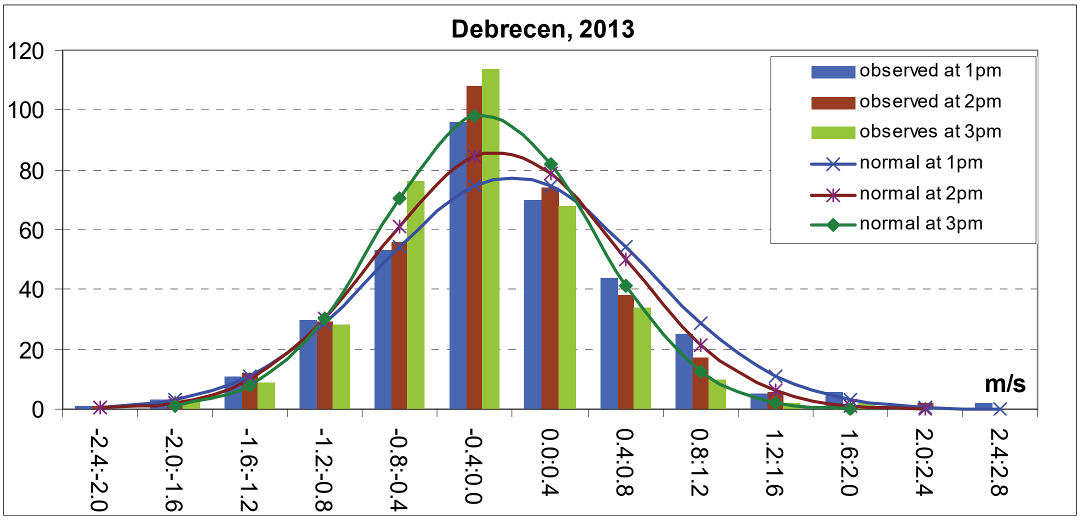

Fig. 8. Frequency distribution (days) of the magnitude of the daily simple estimation errors and their approximation to the normal distribution.

3.3. Relative error of the estimations

We examine the more important statistical properties of the exact relative deviation of the results of the estimates from the known daily average wind speed. For this, we use the absolutely value-free form of Eq.(6) and Eq.(7).

Fig. 9 shows the average monthly relative errors. It can be seen that their value decreases every month as the time of the estimate increases. The average monthly relative error is positive in a total of 7 cases: for all three cases in February and only for the estimates at 1pm and 2pm in November. The primary and secondary maximums, i.e., when the number of overestimates is greater than that of underestimates are found in these months.

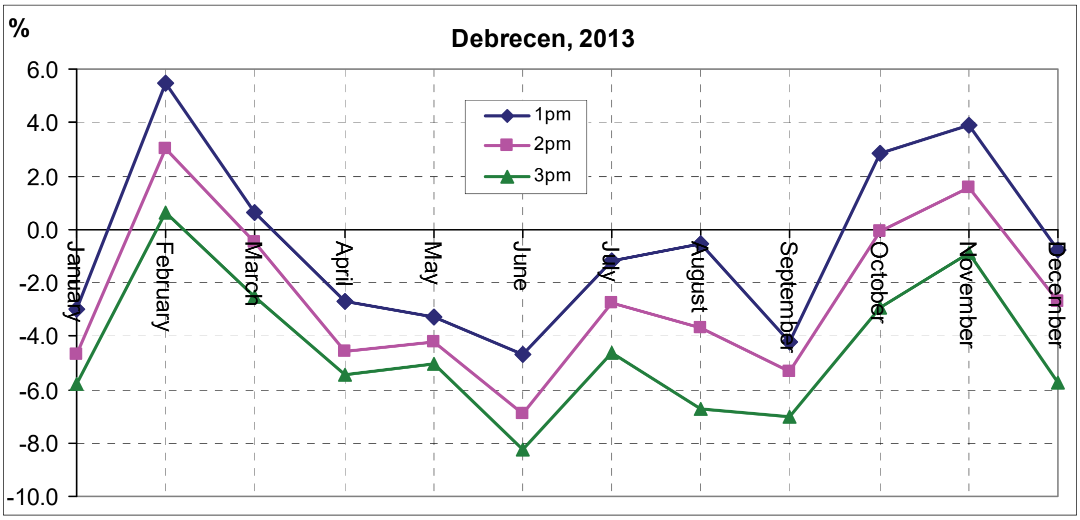

Fig. 9. Annual course of the monthly average relative error of the estimates.

The most important statistical characteristics of the annual process of day-to-day relative error are shown in *Table 5*.

Table 5. The most important statistical characteristics (%) of the annual course of the relative error per day

at	1pm	2pm	3pm
mean	-0.74	-2.68	-4.62
st. deviation	18.56	16.75	14.87
maximum	58.30	47.90	37.50
minimum	-53.10	-53.10	-53.10
range	111.50	101.00	90.60
skewness	-0.03	-0.08	-0.12
kurtosis	0.15	0.20	0.18
mode	0.00	0.00	0.00
median	0.00	-3.00	-5.30

The characteristics of the relative error per day decrease with time except for the minimum value and the mode, which do not change. The comparison of the mean, mode, and median values once again raises the possibility that the studied data originate from a normal distribution. To determine this, the frequency distribution of the magnitude of the studied errors are exemined taking into account the extreme values with classifying them into intervals of 10%.

Fig. 10 shows the observed frequencies and those approached with normal distribution at the three estimation times. According to the χ^2 test, the hypothesis that the frequency distribution of the daily relative estimation errors is also normal is not rejected at 0.05 significance level for the estimates at either hour. This means that relative errors between -10% and 0% are most likely to occur at all three estimation times (1pm, 2pm, and 3pm).

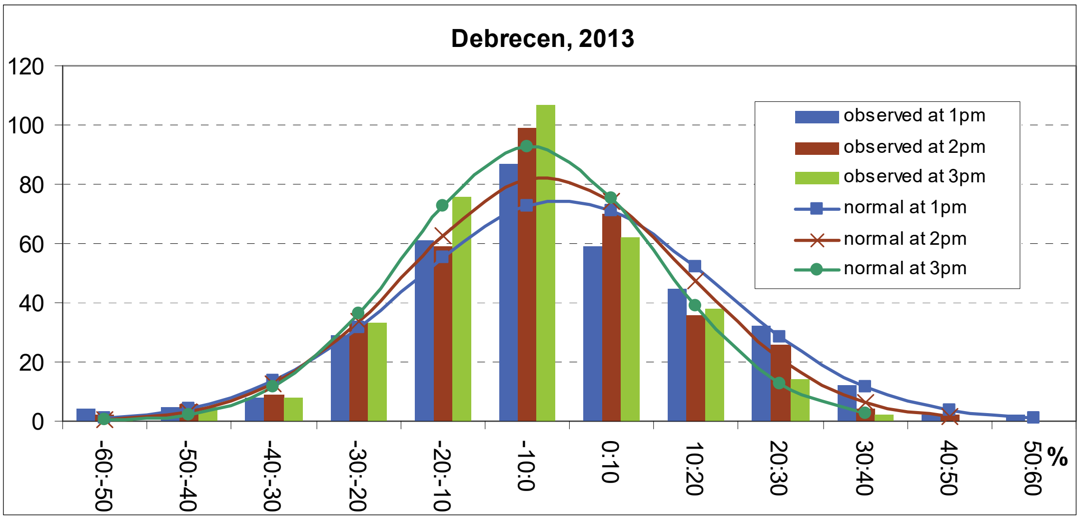

Fig. 10. Frequency distribution (days) of daily relative estimation errors and their approximation to the normal distribution.

Examining the statistical characteristics of the daily simple and relative estimation errors does not give an accurate picture of the difference between estimated and true values. Errors with different signs may, for example, balance each other in the course of averaging, therefore, an average error close to zero may be obtained. The real differences, i.e., the absolute values of these errors, are a more pronounced indicators of the reliability of our model. Therefore, we now examine the absolute value (magnitude) of the daily relative errors, as you see in Eq.(6). Eq.(7) gives the average of these for different periods. From *Fig. 5* it can be concluded that in Debrecen, this error is on average between 15% and 12% at the three selected estimation times in the whole period in the case of the 10-year-long time series used in model construction.

Based on *Fig. 11*, monthly averages are the highest in November and the smallest in May at all three estimation times. In May, August, and December, the average of the three estimates is almost the same, and in the other months – except March – average errors decrease as the estimation times increase.

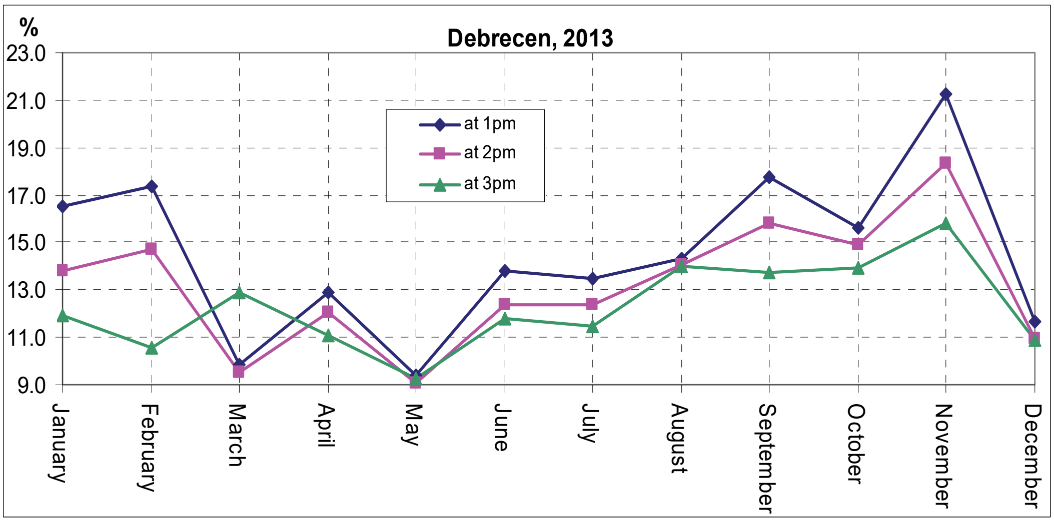

Fig. 11. Monthly averages of the magnitude of the daily relative estimation errors.

Looking at the annual pattern of the magnitude of the daily relative estimation error, its most important statistical characteristics and frequency distribution were determined (*Table 6*). Averages are a very good match to the values shown in *Fig. 5*. To determine the mode, intervals of 5% were used to create frequency distributions shown in the columns in *Fig. 12*. Most data can be categorised into the 0-5% interval at 1pm and 3pm while into the 5–10% interval at 2pm. The centre of these can be considered the value of the mode.

Table 6. The most important statistical characteristics (%) of the annual course of the magnitude of the relative estimation error per day

at	1pm	2pm	3pm
mean	14.68	13.35	12.31
st. deviation	11.36	10.44	9.52
maximum	58.33	53.13	53.13
minimum	0.00	0.00	0.00
skewness	1.06	1.11	1.12
kurtosis	1.17	1.28	1.60
mode	2.50	7.50	2.50
median	12.50	10.53	10.34

It follows from both the minimum and the mode values that the frequencies can be approximated by a monotonous descending theoretical distribution. Exponential and gamma distributions were tested. Despite the fact that the parameter determining the shape of the gamma distribution is in no case smaller than 1, in which case the mode would fall in the 0-5% interval (*Dévényi* and *Gulyás*, 1988), the latter proved to be successful. Thus, the maximum of the theoretical distributions in all three cases is within the 5-10 % interval, not only at the 14-hour estimate (*see Fig. 12*). However, based on the χ^2 test, the hypothesis that the frequency distribution of the magnitude of the daily relative estimation errors has gamma distribution at a significance level of 0.05 is not rejected for any of the hourly estimates. This means that real differences between 0 % and 10% are most likely to occur at all three estimation times.

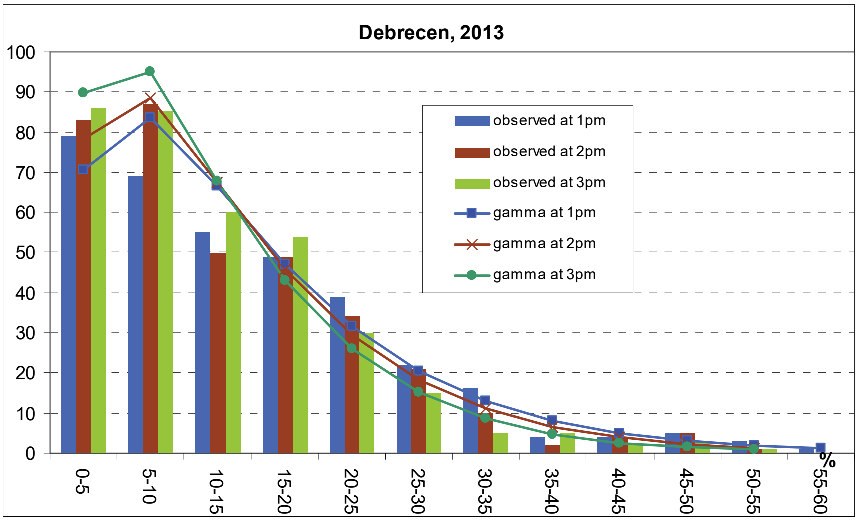

Fig. 12. Frequency distribution (days) of the magnitude of daily relative estimation errors and their approximation to the gamma distribution.

3.4. The estimated present-day and the accurate previous-day daily average wind speed

Based on of above results, if we have decided which estimate to accept as the average wind speed of the present day, this can be compared with the known average wind speed of the previous day, it can be decided whether the average wind speed of the present day and with it the average daily wind power decreased or increased compared to the previous day.

Considering the difference between the average wind speeds estimated for the present day and known for the previous day, if the difference is negative, the average wind speed of the present day decreases compared to that of the previous day. If the sign of this difference is compared with the sign of the difference in real average wind speeds, very valuable information can be obtained regarding the reliability of the model.

Let us have, as an example, the average wind speed of the present day, estimated at 1pm, $\bar{v}_{te}$ =4.3 m/s, and that of the previous day, $\bar{v}_{yr}$ =5.3 m/s. Their difference is negative, therefore, the average wind speed of the present day is expected to be smaller than 5.3 m/s. At the end of the day, the average wind speed of the present day is found to be $\bar{v}_{tr}$ =3.7 m/s, which means that the estimate was correct. Therefore, reliability of the model can also be tested so that the sign of the $\bar{v}_{te} - \bar{v}_{yr}$ and $\bar{v}_{tr} - \bar{v}_{yr}$ differences are compared. Their similarity indicates that the estimation is correct.

Similarity of the signs is tested monthly, taking into account the missing days as $\bar{v}_{yr}$ is missing on the following days. The number of days when this was the case – i.e., sign($\bar{v}_{te}$–$\bar{v}_{yr}$)=sign($\bar{v}_{tr}$– $\bar{v}_{yr}$) – compared to the number of days can be considered in the model (in %) are presented in *Fig. 13*.

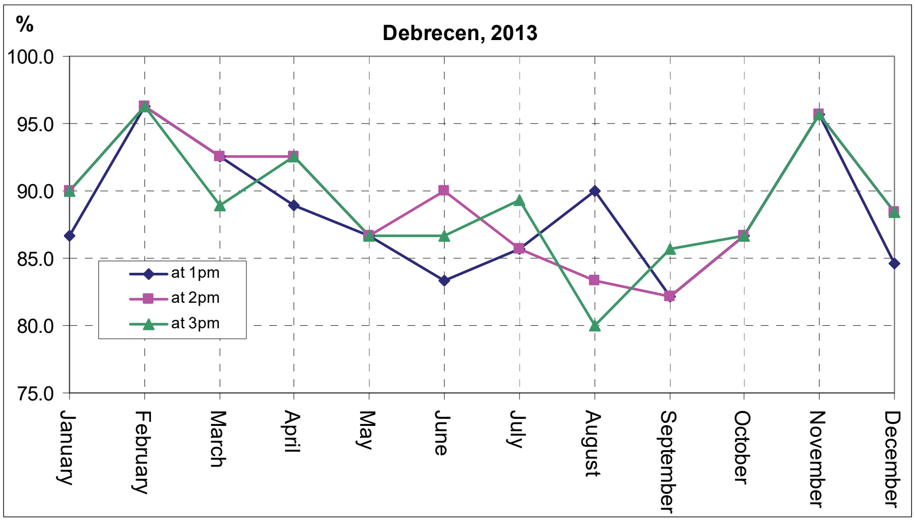

Fig. 13. The identical signs of the difference between the estimated current and yesterday's real, and today's real and yesterday's real average wind speeds per month in relation to the number of days to be taken into account (in %).

According to the chart, the maximum of the proportions is set in February at 96.3% for all three estimates. This is also the case in November with a secondary maximum of 95.7%. The minimum value is 82.1% in September at 1pm and 2pm and 80.0% at 3pm in August. In other words, the sign of the differences examined is between 80.0% and 96.3%. Our assumption, that these rates do not decrease as the estimate date increases, is true with the exception of March, June, and August.

For the whole year, these rates are 88.6, 89.5, and 89.2, i.e., with a slight maximum at 2pm. On average, therefore, there is a nearly 90% probability that if the average wind speed of the present day is smaller or greater than the real average wind speed of the previous day, then the real wind speed of the present day will change accordingly by the end of the day.

3.5. *Change in the average wind speed of the next day compared to that of the present day based on the estimated present-day average wind speed*

In two previous articles (*Tar* and *Lázár*, 2018; *Tar*, 2021), the average daily wind speeds observed at different stations were transformed to 10 m. In addition to comparability, this was also justified by the fac,t that the height of the anemometer at certain stations changed during the studied period, and therefore, a reference level was also required. The main conclusion of the detailed analysis carried out is that if the average wind speed of the present day is less than the average speed of the category (year, season, macrosynoptic situation group) over many years, then the increase of the average wind speed of the following day is 1.4–2.3 times more likely, on average, 1.9 times more likely than its decrease. However, if the average daily wind speed is greater than the average speed of the category, the decrease of the average wind speed of the following day is 1.6–5.2 times, on average, 2.4 times more likely than its increase.

The change in the average wind speed of the next day compared to that of the present day is determined from the average wind speeds of the present day estimated as detailed above using the annual 30m SODAR data of 2013. According to the above, these observed and estimated daily average wind speeds should be transformed to 10 m. This was performed using the so-called WMO formula (*Mezősi* and *Simon*, 1981), according to which the values at the 10 m height is about 80 % of those at 30 m. The average wind speed at 10 m during the modeled (studied) period (1991–2000) was 2.8 m/s.

The procedure of the estimation is detailed in *Table 7*. Here $[v]_e$ represents the average wind speed at 10 m estimated at 1pm of the given day. The next column shows the difference from the average wind speed of the studied category, i.e., the total period (2.8 m/s). If this is negative or 0, the average wind speed of the next day, based on the above, is most likely to increase or not to decrease (I), otherwise it will decrease (D). Based on the data in the I/D_e column, we estimate that in 32.3% of the days of the month, the average wind speed of the next day will increase compared to that of the previous day. This can be controlled, as daily averages at 10m are available for the test period. These are indicated by [v] in the table, and Δv indicates their differences on the following days. The increase is now the $\Delta v \geq 0$. The data in column I/D show that the true frequency of increase (I) is 58.1%, which was therefore underestimated by about 25%. If the same event (I or D) occurs with an estimation and observation on a given day, it can be considered a good estimate. The total number of good estimates (II+DD) in the table is 17, i.e., in January the estimate at 1pm and the observation are in accordance with each other in the case of 54.8% of the days.

Table 7. Illustration of estimation process (bold-italic letters indicate good estimate)

	estimation at 1pm			observed		
January	**$[v]_e$**	**$[v]_e$-2.8**	**I/D_e**	**[v]**	**Δv**	**I/D**
1.	0.9	-1.9	***I***	0.9	0.2	***I***
2.	0.6	-2.2	***I***	1.1	3.1	***I***
3.	3.1	0.3	***D***	4.3	-1.3	***D***
4.	3.5	0.7	D	3.0	1.7	I
5.	3.8	1.0	***D***	4.7	-1.7	***D***
6.	3.5	0.7	D	3.0	2.7	I
7.	6.6	3.8	***D***	5.6	-2.9	***D***
8.	3.2	0.4	***D***	2.7	-0.8	***D***
9.	2.5	-0.3	***I***	1.9	0.3	***I***
10.	2.1	-0.7	***I***	2.3	1.1	***I***
11.	3.1	0.3	***D***	3.4	-0.6	***D***
12.	2.4	-0.4	***I***	2.8	0.3	***I***
13.	2.3	-0.5	***I***	3.1	0.4	***I***
14.	3.8	1.0	***D***	3.5	-0.8	***D***
15.	2.8	0.0	D	2.7	0.2	I
16.	3.6	0.8	D	2.9	1.9	I
17.	4.7	1.9	D	4.8	1.4	I
18.	6.7	3.9	***D***	6.2	-3.4	***D***
19.	2.7	-0.1	I	2.8	-0.1	D
20.	2.9	0.1	D	2.7	0.2	I
21.	2.9	0.1	D	2.9	0.6	I
22.	3.1	0.3	***D***	3.5	-0.2	***D***
23.	3.6	0.8	D	3.2	1.7	I
24.	3.4	0.6	D	4.9	2.3	I
25.	7.7	4.9	***D***	7.2	-3.6	***D***
26.	4.3	1.5	***D***	3.6	-1.4	***D***
27.	1.7	-1.1	N	2.2	-0.7	D
28.	1.8	-1.0	***I***	1.4	2.7	***I***
29.	2.3	-0.5	I	4.1	-0.3	C
30.	3.1	0.3	D	3.8	1.4	I
31.	4.3	1.5	D	5.2	1.6	I

The analysis detailed above was carried out monthly for all three estimate times. Therefore, the reality of the assumed periods (half-year, seasonal) and the best estimate time can be identified. Finally, results for the whole year are also provided.

Fig. 14 shows the monthly frequency of the occurrence of event N in the percentage of the applicable days in the given month.

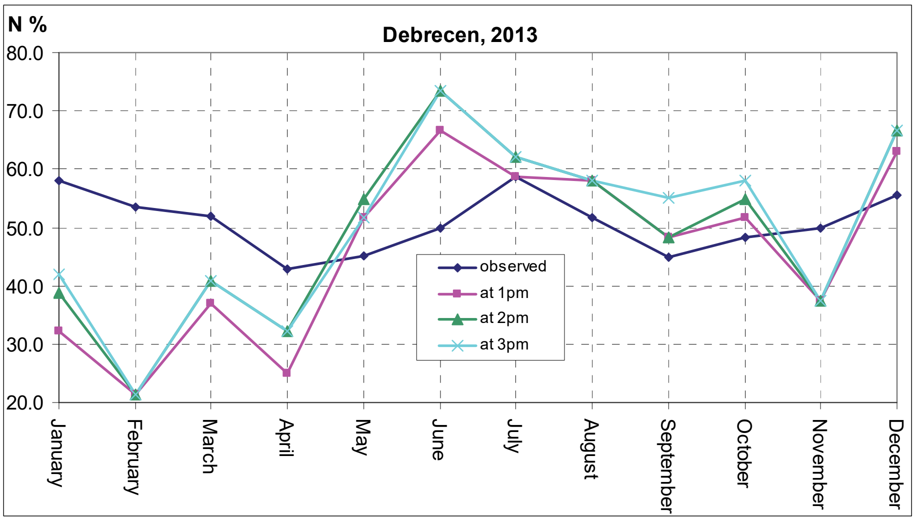

Fig. 14. Observed and estimated monthly frequencies of occurrence of event N in % of usable days at the three time points.

The figure shows that the difference between the estimates at the three different times is unlikely to be significant. On average, the estimates at 3pm and 2pm differ from each other the least, by 0.8%, and those at 3pm and 1pm differ the most, by 4.0%. However, all three estimates approximate the observed frequency with statistically reasonable accuracy only during the July-October period.

The regularity in the annual pattern of the curve showing the observed frequency suggests the presence of a real period. According to the period analysis (*Dobosi* and *Felméry*, 1971; *Tar* and *Kircsi*, 2001; *Tar et al.*, 2002; *Matyasovszky*, 2002; *Tar*, 2007, 2008ab), the wave with a half-year period has a real, non-random annual pattern. This means that the probability of the occurrence of event N during the winter and summer months is significantly higher than in the other two (transitional) seasons (see *Fig. 15*).

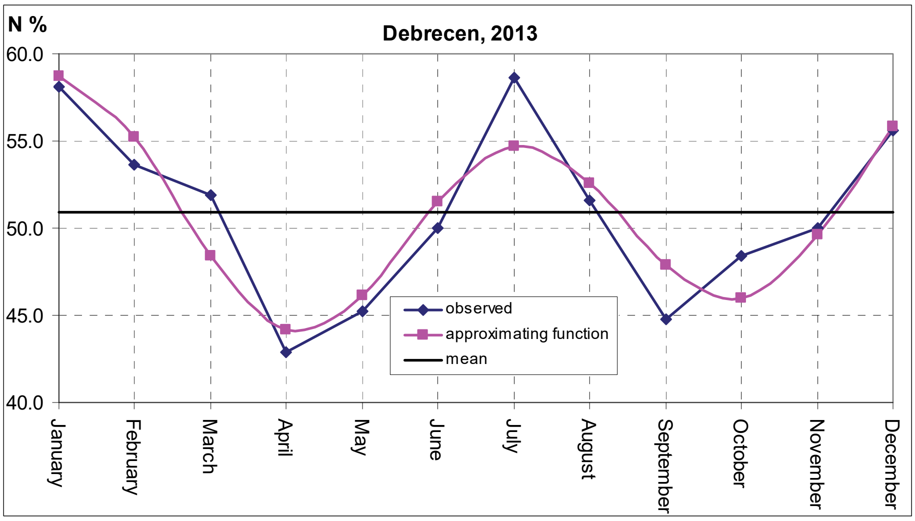

Fig. 15. The annual course of the observed frequency of event N and the trigonometric function with a realistic period of half a year approaching it.

Regarding the whole year, on the basis of the observed (real) values, the average daily wind speed decreased in 49.1% of all days compared to the previous day and increased (not decreased) in 50.9%. The probability of the two events can therefore be considered roughly equal in the studied year. The values calculated from the estimates at 3pm are the closest to the values of 49.7% and 50.3%. Regarding the other two estimate times, the frequency of event C is 7.5% and 1.2% higher than that of event N.

Monthly frequencies of good estimates are shown in *Fig. 16* as a % of the days that can be used in a given month.

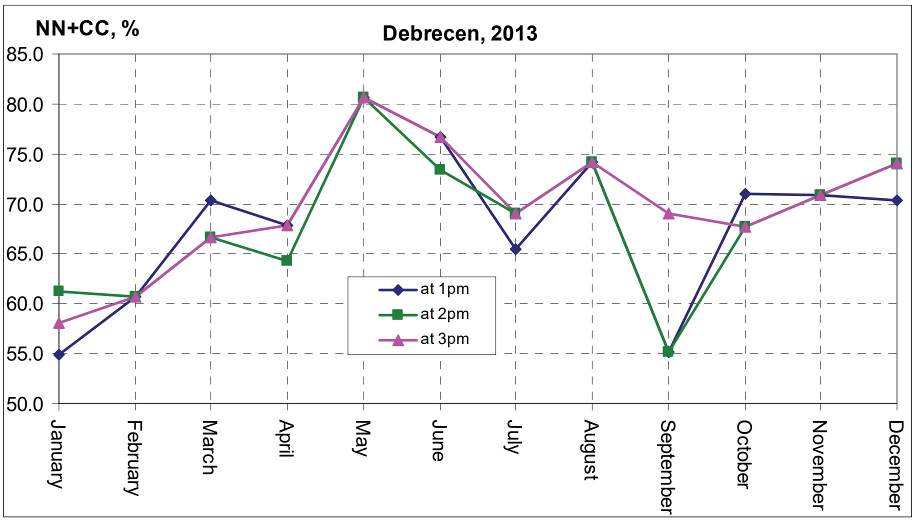

Fig. 16. Monthly frequencies of coincidence of estimated and observed N and C events.

The frequency of good estimates ranges from 80.6% (May, all three times) to 54.8% (January, 1pm). The curves run well together except for September, when there is a nearly 14% difference between the frequencies at 3pm and also at 1pm and 2pm. On an annual average, however, the frequency of good estimates for 1pm and 2pm is only 1.5% behind the estimates at 3pm, which is 69.7%.

The September estimate was also carried out with the autumn estimation parameters: with 1.01, 1.03 and 1.05 [R_i] values and a seasonal average of 2.5 m/s. However, the number of good estimates decreased at 2pm and 3pm compared to the values shown in *Fig. 16*.

4. Summary

According to our previous study, the day-to-day change of the average daily wind speed has an appreciable stochastic relationship with the average wind speed of the previous day. The conclusion based on the regression function is that if the average wind speed of the present day is less than the average speed of the studied period (e.g., year, season), average increase in the average wind speed of the next day is nearly twice as likely as the decrease. However, if the average daily wind speed is greater than the average speed of the category, decrease in the average wind speed of the next day is, on average, nearly two and a half times more likely than its increase. Thus, knowing the average speed of the present day, makes it possible to estimate the sign of a change for the next day, which can help to prepare a timetable. However, in order to perform this estimate, the average wind speed of the present day has to be estimated as well at the time of day that can be used for the preparation of the timetable. The sliding average model was established to face this problem.

After the detailed analysis of the daily pattern of the annual and seasonal average hourly sliding averages ([R_i], i=1,2,...,24) produced from the ten-year hourly wind speeds of five Hungarian meteorological stations, the average daily wind speed was estimated from each value of the above using a simple ratio, and the relative error of the estimate was also determined.

The average hourly relative error ([E_i]) decreases rapidly as the estimate time approaches the end of the day. The average hourly decrease is the highest in Szombathely and the lowest at Kékestető regarding absolute values, and the order between them is Debrecen, Szeged, Budapest in all categories (years, seasons). As a result of rapid decreases, the values of [E_i] fall below 20% after 1pm in all cases.

For the testing of the model, the annual SODAR data measured at 30 m height in Debrecen in 2013 were used. Daily average wind speed is estimated at 1pm, 2pm, and 3pm using the average relative sliding averages ([R_i]) determined for the whole period (year) monthly. The estimate is considered accurate if its difference from the real daily average is 0.0 to one decimal point. Underestimation and overestimation mean that the difference is negative or positive, respectively.

The maximum of the monthly relative frequency of accurate estimates is divided between 1pm and 3pm in almost 50%–50% of the cases. The difference between the monthly ratio of underestimates and overestimates is negative in February and November (i.e., the number of overestimates is higher in these months), but with the exception of November, this difference is the largest for the 3pm estimate. Therefore, the number of underestimates increased as the estimate time increases in the present case. The evaluation of the full-year results of the estimates also shows this.

The sign estimation error is the difference between the estimated and the real daily average wind speed. The characteristics of the estimation error decrease over time, except for the minimum value and mode, and the minimum increases. Modes can be considered equal. This can be decided by the frequency distribution of the magnitude of errors. According to this, errors occur with the highest frequency within the interval (-0.4:0.0) m/s at all three times, the same way as averages and medians. Other parameters also suggest the possibility of approaching with normal distribution. According to the χ^2 test, the hypothesis that the frequency distribution of the magnitude of estimation errors has normal distribution at a significance level of 0.05 is not rejected for either estimates. This means that deviations between -0.4 and 0.0 m/s are most likely to occur at all three estimation times.

Then the main statistical properties of the relative differences of the results of the estimates in % relative to the exact average daily wind speed are known at the time of testing. The comparison of the mean, mode, and median values raises again the possibility that the studied samples are from a normal distribution. To decide this, the frequency distribution of the magnitude of the studied errors, with an interval of 10% taking into account the extreme values is exemined. According to the χ^2 test, the hypothesis that the frequency distribution of the magnitude of estimation errors has normal distribution at a significance level of 0.05, is not rejected for either estimates. This means that relative errors between -10% and 0m/s are most likely to occur at all three estimation times.

Examining the statistical characteristics of daily simple and relative estimation errors gives no accurate picture of the difference between estimated and real values. Real differences, i.e., the absolute values of these errors, are more pronounced indicators of the reliability of our model. Therefore, the monthly characteristics of the absolute value of daily relative errors were also examined. In Debrecen, at the three selected estimates, this error is on average between 15% and 12% in Debrecen at the three selected estimates for the 10-year-long time series used in model construction in the whole period. With this, the monthly averages of the values calculated here are in a very good match. Intervals of 5% were used to create frequency distributions for determining the mode. Most elements of the sample classify into the interval 0–5% at 1pm and 3pm, however, at 2pm most data belong to the 5–10% interval. Frequencies can therefore be approximated by a monotonous descending theoretical distribution. Based on the

χ^2 test, the hypothesis that the frequency distribution of the magnitude of the daily relative estimation errors has gamma distribution is not rejected at a significance level of 0.05 for any estimation time. This means that real differences between 0% and 10% are most likely to occur at all three estimation times.

Comparing the present-day average wind speed estimate with the known, real, accurate average wind speed of the previous day, it can be assumed whether the average wind speed, and with it, the average daily wind power of the present day will decrease or increase compared to those of the previous day. Let us take the difference between the estimated present-day and the known average wind speed of the previous day. If this difference is negative, then that of the present day will decrease compared to that of the previous day. If the sign of this difference is compared to the sign of the difference between the real, accurate average speeds, their matching indicates correct estimation. The maximum of correct estimates occurs in February, which is 96.3% for all three estimates. This is also the case in November with a secondary maximum of 95.7%. The minimum value is 82.1% in September at 1pm and 2pm, while at 3pm, it is 80.0% in August. In other words, the sign of the differences examined is between 80.0% and 96.3%. For the whole year, these rates are 88.6, 89.5, and 89.2, i.e., with a slight maximum at 2pm. On average, therefore, there is a nearly 90% probability that the average wind speed of the present day is less than or greater than the real average wind speed of the previous day, then the real wind speed of the present day will change accordingly by the end of the day.

Finally, based on the results of the testing the model was applied to solve the original problem. To do this, the estimated average wind speed of the present day had to be compared with the long-term average wind speed of the studied period (category). As described at the beginning of this section, if the estimated average wind speed of the present day is less than or greater than the average speed of the period, which is currently 2.8 m/s, then the increase (N) or decrease (C) of the average wind speed of the next day is more likely. The number and proportion of good estimates can be determined by comparing estimates with real average daily wind speeds.

The analyses were carried out monthly and throughout the year for all three estimation times. The monthly occurrences of event N indicate, that the difference between the three estimates is unlikely to be significant. On average, the estimates at 3pm and 2pm differ the least, by 0.8%, and the estimates at 3pm and 1pm differ the most, by 4.0%. However, all three estimates approximate the observed frequency with statistically reasonable accuracy only during the July-October period. For the whole year, on the basis of the observed (real) values, the average daily wind speed decreased in 49.1% of all days compared to the previous day and increased (not decreased) in 50.9%. The probability of the two events can therefore be considered to be equal with a good approximation in the studied year. The values calculated from the estimates at 3pm approximate best the above values: 49.7% and 50.3%. Regarding the other two estimates, the frequency of event C is 7.5% and 1.2% higher than that of event N.

Events N and C are assigned to the present day and show that the average wind speed for the next day will increase or decrease. The time series of these events can be produced by comparing the estimated average daily wind speeds with the long-term average. A good estimate is obtained when the event assigned to the given day based on the estimation is the same as the event that can be determined from the real (known at testing) daily average wind speeds. These are events NN and CC, thus the number of good estimates is NN+CC. Curves in the annual pattern look quite the same except for September, when there is a nearly 14% difference between the frequencies related to 3pm, 1pm, and 2pm. On an annual average, however, the frequency of good estimates at 1pm and 2pm is only 1.5% behind those at 3pm, the latter is 69.7%.

From the detailed analyses above, it can be seen that by merging the model describing the daily average wind speed with the sliding average model, more information about the wind climate in Hungary can be revealed, in addition to the hopes that it will help to prepare the timetable. For a given wind power plant, the method is considerably easier to apply, since the long-term wind speeds measured there can produce the characteristics needed to operate both models.

Acknowledgements: The authors would like to express their thanks to the Hungarian Meteorological Service for providing data for the analysis.

References

Aggarwal, S.K. and *Meenu Gupta*, 2013: Wind Power Forecasting: A Review of Statistical Models. *Int. J. Energy Sci.3*(1).

Armstrong, J.S. and *Collopy, F.,* 1992: Error measures for generalizing about forecasting methods: Empirical comparisons. *Int. J. Forecast 8*, 69–80.

Bremnes, J.B., Villanger, F., and *AS, K.V.,* 2002: Probabilistic forecasts for daily power production. Proceedings of the Global Wind Power Conference, Paris.

Dévényi, D. and *Gulyás, O.,* 1988: Matematika statisztikai módszerek a meteorológiában. Tankönyvkiadó, Budapest. (in Hungarian)

Dobosi, Z. and *Felméry*, *L.,* 1971: Klimatológia. Egyetemi jegyzet, Tankönyvkiadó, Budapest. (in Hungarian)

Kavasseri, R.G. and *Seetharaman, K.,* 2009: Day-ahead wind speed forecasting using f-ARIMA models. *Renew. Energy 34,* 1388–1393. *https://doi.org/10.1016/j.renene.2008.09.006*

Károssy, Cs., 1993: A Péczely-féle makroszinoptikus tipizálás és a helyzetek katalógusa (1951–1992). In (Ed.: *Nowinszky L.*) A fénycsapdás rovargyűjtést módosító abiotikus tényezők. I. kötet, OSKAR Kiadó, Szombathely,113–126. (in Hungarian)

Károssy, Cs., 1998: Péczely's classification of macrosynoptic types and catalogue of weather situations (1992–1997). In (Ed.: *Nowinszky L.*) Light trapping of insects influenced by abiotic factors. Part II, Savaria University Press, 117–130.

Károssy, Cs., 2001: Characterisation and catalogue of the Péczely's macrosynoptic weather types (1996–2000). In (Ed.: *Nowinszky L.*) Light trapping of insects influenced by abiotic factors. Part III. Savaria University Press, 75–86.

Matyasovszky I., 2002: Statisztikus klimatológia. Idősorok elemzése. ELTE Eötvös Kiadó, Budapest. (in Hungarian)

Mezősi, M. and *Simon, A.,* 1981: A meteorológiai szélmérés elmélete és gyakorlata. Meteorológiai Tanulmányok 36. (in Hungarian)

Olaofe, Z.O., Folly, K.A., 2012: Statistical Analysis of the Wind Resources at Darling for Energy Production. *Int. J. Renew. Energ. Res. 2*, 250–261.

Péczely, Gy., 1961: Magyarország makroszinoptikus helyzeteinek éghajlati jellemzése. Az Országos Meteorológiai Intézet Kisebb Kiadványai 32. (in Hungarian)

Shukur, O.B. and *Lee, M.H.,* 2015*:* Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA. *Renew. Energy 76,* 637–647. *https://doi.org/10.1016/j.renene.2014.11.084*

Tar, K., 1990: Statistical Investigation on the 130-year Time Series of precipitation in Debrecen. Climatic Change in the Historical and Instrumental Periods. Masaryk University-Brno, 275–279.

Tar, K., 1993: Investigation of the Time Series of the Monthly Relative Sums of Precipitation. Early Meteorological Instrumental Records in Europe. Uniwersytet Jagiellonski, Kraków, 183–191.

Tar, K., 1995a: A havi relatív csapadékösszegek idősorának tulajdonságai Magyarországon. Berényi Dénes professzor születésének 95. évfordulója tiszteletére tudományos emlékülés előadásai. KLTE Debrecen, 158–165. (in Hungarian)

Tar, K., 1995b: Investigation of the time series of the monthly relative sums of precipitation in Hungary. Proceeding of Conference in Atmospheric Physics and Dynamics in the Analysis and Prognosis of Precipitation Fields, Rome, 405–410.

Tar, K. and *Kircsi, A*., 2001: Kísérlet a szélenergia statisztikai becslésére. Szélenergia konferencia, Gödöllő, 28–34. (in Hungarian)

Tar, K., Kircsi, A., and *Szegedi, S.*, 2001: A possible statistical estimation of wind energy. Proceedings of the European Wind Energy Conference, Copenhagen, Denmark, 886–889.

Tar, K., Kircsi, A., and *Vágvölgyi, S*., 2002: Temporal changes of wind energy in connection with the climatic change. Proceedings of the Global Windpower Conference and Exhibition, Paris, France, 2-5 April, CD-ROM.

Tar, K., 2004: Becslési módszerek a magyarországi szélenergia potenciál meghatározására. *Magyar Energetika 12*(4), 37–48. (in Hungarian)

Tar, K., 2007: Diurnal course of potential wind power with respect to the synoptic situation. *Időjárás 111,* 261–279.

Tar, K., *Maghiar, T., Bondor, K.* and *Szegedi, S*., 2007: Statistical estimation of diurnal average potential windpower. Proceedings of the 9th International Conference on Engineering of Modern Electric System, Oradea, Romania, 86–90.

Tar, K., 2008a: Az időjárási helyzetek szélenergiájáról. Tanulmánykötet Dr. Gööz Lajos professzor 80. születésnapjára. Nyíregyháza, 267–276. (in Hungarian)

Tar, K. 2008b: Energetic characterization of near surface windfield in Hungary. *Renew. Sustain. Energ. Rev. 12*, 250–264.

Tar, K. and *Szegedi, S*., 2011: A statistical model for estimating electricity produced by wind energy. *Renew. Energ. 36,* 823–828. DOI: 10.1016/j.renene.2010.06.032.

Tar, K. and *Lázár, I.*, 2018: Statistical structure of day by day alteration of daily average wind speeds. *Időjárás 122*, 285–304.

Tar, K., 2019: Statisztikai módszer a napi átlagos szélsebesség napközbeni becslésére. Tiszteletkötet Puskás János 65. születésnapjára. Szombathely, 115–127. (in Hungarian)

Tar, K., 2021: Az átlagos szélsebesség napi változásának statisztikai becslése. *Légkör 66*(1)*,* 27–32. (in Hungarian)