MULTIVARIATE ANALYSIS OF TWO-YEAR RADON CONTINUOUS MONITORING IN GROUND LEVEL LABORATORY IN THE INSTITUTE OF PHYSICS BELGRADE

by

Dimitrije M. MALETIĆ [1], Radomir M. BANJANAC [1], Dejan R. JOKOVIĆ [1], Aleksandar L. DRAGIĆ [1], Nikola B. VESELINOVIĆ [1], Mihailo R. SAVIĆ [1], Zoran R. MIJIĆ [1], Vladimir I. UDOVIČIĆ [1*], Svetlana D. ŽIVKOVIĆ-RADETA [2], and Jelena V. UDOVIČIĆ [2]

[1] Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia
[2] Faculty of Applied Ecology Futura, Belgrade Metropolitan University, Belgrade, Serbia

Scientific paper
https://doi.org/10.2298/NTRP2304273M

Multivariate classification and regression analysis of multiple meteorological variables and indoor radon activity concentration in Ground Level Laboratory in the Institute of Physics Belgrade, was performed and discussed. Meteorological variables used in this analysis were from radon active device, nearby meteorological station and finally from Global Data Assimilation System. Single variate analysis has identified variables with greatest value of Pearson's correlation coefficient with radon activity concentration and also, variables with greatest separation of events with increased radon activity concentration of over 200 Bqm^{-3} and of events with radon level below this value. This initial analysis is showing the expected behavior of radon concentration with meteorological variables, with emphasis on data periods with or without air conditioning and with emphasis on indoor water vapor pressure, which was, in our previous research, identified as important variable in analysis of radon variability. This single variate analysis, including all data, proved that Global Data Assimilation System data could be used as a good enough approximate replacement for meteorological data from nearby meteorological station for multivariate analysis. Variable importance of Boosted Decision Trees with Gradient boosting multivariate analysis method are shown for all three periods and most important variables were discussed. Multivariate regression analysis gave good results, and can be useful to better tune the multivariate analysis methods.

Key words: continuous radon monitoring, multivariate analysis, Global Data Assimilation System, meteorological station

INTRODUCTION

Primarily, radon problem presents a health hazard [1]. The research of the dynamics of radon in various environments, living or working places, is of great importance in terms of protection against ionizing radiation and in designing of measures for its reduction. In the Low-Background Laboratory for Nuclear Physics extensive research on various radon fields has been done in the past, especially radon monitoring in the special designed low-background underground and ground level laboratory, with the aim of investigating the rare nuclear processes [2]. Besides radon monitoring in the laboratory, we work on several research topics regarding radon: using multivariate classification and regression methods, as developed for data analysis in high-energy physics [3], to study connection of climate variables and variations of radon concentrations, modelling of the indoor radon behaviour and national indoor radon mapping [4], taking interest in similar indoor radon mapping analysis in Montenegro [5], or by research of radon variability in a single dwelling [6], using advanced analysis tools, or performing continuous measurements in multi-store building [7] or laboratory space [8]. Indoor radon variability depends on many variables. Soil content, and building characteristics are very important. In case of researching of indoor radon variability, meteorological effects become the most important ones. With recent experiences with lowering the limits of indoor radon level, both in dwellings and working places, and the demand for decrease of public radon exposure, the need for more detailed knowledge on radon variability is increasing. Besides a possibility for improvement of mitigation

* Corresponding author, e-mail: udovicic@ipb.ac.rs

techniques, we could look into creating online warning pages, like we already have, for example, for UV radiation. These online warning pages, with information on radon concentration variations, could be interesting to people living in dwellings or working spaces with previously known radon problem, or dwellings with radon activity concentration close to 200 Bqm^{-3} limit. These online warnings, could indicate a call for some temporary measures like starting of increased ventilation or reducing exposure. Local radon warning pages could be based on local meteorological station, but for larger regions, meteorological modeled data like Global Data Assimilation System (GDAS) could be used. In this paper we were looking into the possibility of using GDAS data in prediction of indoor radon variability, by jointly looking into GDAS and nearby meteorological station, and compare the results.

DATA PREPARATION AND SELECTION

The radon continuous monitoring in ground level laboratory was performed with active device RadonEye Plus2 with time sampling of one hour. The device recorded variables: Rn-activity, indoor temperature and indoor humidity. The radon the measurement was done from November 2020 to November 2022. After looking into indoor temperature data, we decided to do three analysis, one with using all the data samples (whole period of measurement's), second using only data when air conditioning (AC) was operating, and third sample used for analysis was for periods when air conditioning was OFF (noAC).

Meteorological station located in Institute of Physics Belgrade yard, and maintained by Environmental Physics Laboratory [8], has being recording variables at 5 minute interval, and hourly values are used for this analysis. Variables are named by adding prefix outside; outside-cloudbase, outside-dew point, outside-humidity, outside-temp, outside-pressure and outside-rain.

The US National Centers for Environmental Prediction (NCEP) runs a series of computer analyses and forecasts operationally. One of the operational systems is the GDAS. At National Oceanic and Atmospheric Administration's (NOAA) Air Resources Laboratory (ARL), NCEP model output is used for air quality transport and dispersion modeling. The ARL archives GDAS output which contains basic fields, such as the temperature, pressure and humidity. Those GDAS data are very interesting since they are widely used by weather forecast groups worldwide, and our idea is that if we could use this freely accessed and frequently updated database, we could improve forecasting of some kind of *relative* indoor radon concentrations, and indicate by result of automatic online MVA regression analysis when to expect increased indoor radon concentrations based on meteorological variables. Because MVA methods are rather robust, and we wanted to see which, if any of GDAS variables are suited for our purpose, we included most of variables in our analysis. The GDAS1 data is available for integer values of latitude and longitude, so, for all variables', each data point was firstly 2-D linearly interpolated using variables' values on four integer latitudes and longitudes, surrounding latitude and longitude of our laboratory. The GDAS1 data is available for every three hours, so linear interpolation of each variable's data point was made in order that we can use hourly data. The GDAS1 variables used in our analysis can be identified as ones with prefix GDAS1; GDAS1-CAPE (convective available potential energy), GDAS1-CINH (convective inhibition), GDAS1-CPP6 (accumulated convective precipitation), GDAS1-CRAI (categorical rain), GDAS1-DSWF (downward short wave radiation flux), GDAS1-HCLD (high cloud cover), GDAS1-LCLD (low cloud cover), GDAS1-LHTF (latent heat net flux at surface), GDAS1-LIB4 (best 4-layer lifted index), GDAS1-LISD (standard lifted index), GDAS1-MCLD (middle cloud cover), GDAS1-PBLH (planetary boundary layer height), GDAS1-PRSS (pressure at surface), GDAS1-RH2M (relative humidity at 2m AGL), GDAS1-SHGT (geopotential height), GDAS1-SHTF (sensible heat net flux at surface), GDAS1-SOLM (volumetric soil moisture content), GDAS1-T02M (temperature at 2m AGL), GDAS1-TCLD (total cloud cover), GDAS1-TMPS (temperature at surface), GDAS1-TPP6 (accumulated precipitation), GDAS1-mofi-e (momentum flux intensity), GDAS1-mofd-e (momentum flux direction). In this analysis using GDAS data, we also could indicate if variables measured by local meteorological station do not differ too much from GDAS modeled and interpolated ones, that GDAS variables could be used in this kind of MVA analysis.

We included previously found interesting variable in radon research [6] and that is water vapor pressure in outdoor and indoor air, as well as the difference of the two. In order to calculate the water vapor pressure in air, we need to calculate the value of the saturation water vapor pressure

$$\mathrm{es}(T) = 0.6108 \cdot \mathrm{e}^{\frac{17.27 \cdot T}{T + 237.3}} \tag{1}$$

In addition, the slope of the relationship between the saturation water vapor pressure (es [kPa]) and the air temperature T [°C], is given in [9, 10], so including the slope, we get new formula for the saturation water vapor pressure

$$es(T) = \frac{4098 \cdot \left(0.6108 \cdot \mathrm{e}^{\frac{17.27 \cdot T}{T + 237.3}} \right)}{(T + 237.3)^2} \tag{2}$$

and since the formula used to calculate the relative humidity is

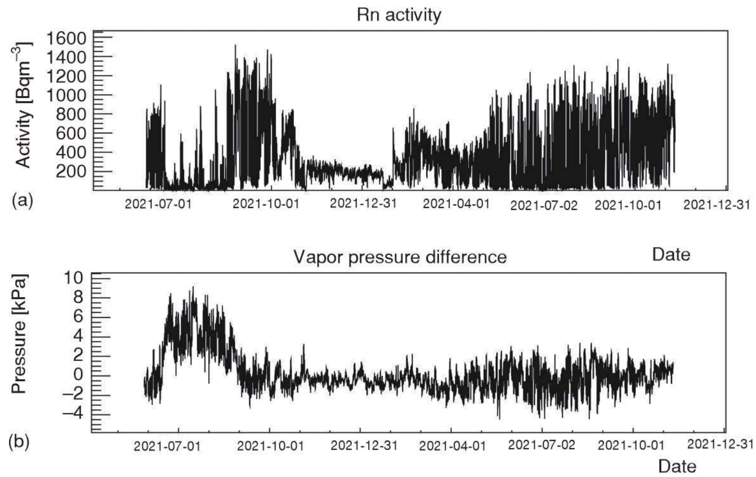


Figure 1. The Rn activity indoor (a) and vapor pressure difference of outdoor and indoor (b). Note that with much greater outdoor water vapor pressure than indoor, comes influx of radon-free water vapor, and that results in significant decrease of indoor Rn activity

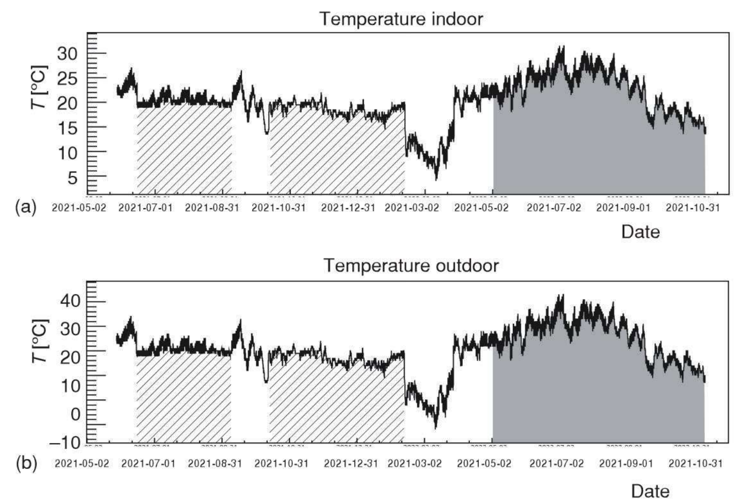


Figure 2. Indoor temperature (a) and outdoor temperature (b) is shown. Indoor temperature which was used for analysis when air conditioning (AC) was on, is indicated in two line pattern areas, while gray shaded interval indicates period when air conditioning was off (noAC)

$$\mathrm{RH} = \frac{\text{vapor pressure}}{\mathrm{es}(T)} \tag{3}$$

we get the formula to calculate the vapor pressure in air

$$\text{vapor pressure}(T, \mathrm{RH}) = \mathrm{RH} \cdot \frac{4098 \cdot \left(0.6108 \cdot \mathrm{e}^{\frac{17.27 \cdot T}{T + 237.3}} \right)}{(T + 237.3)^2} \tag{4}$$

Using this formula, we calculate four variables: indoor-vapor-press (vapor pressure from indoor-temperature and indoor-humidity data), outside-vapor-press (vapor pressure from outdoor outside-humidity, outside-temp data), diff-vapor-press (vapor pressure difference of outdoor and indoor) and gdas1-vapor-press (vapor pressure from GDAS1-T02M, GDAS1-RH2M data). On the bottom of fig. 1 the vapor pressure difference is shown, and it can be clearly seen that if the outer vapor pressure is much higher than the indoor vapor pressure, the indoor radon activity is lower fig. 1(a).

Out of two years of data taking, after merging all the data together to form a single hourly event with all the variables measured at that time, the number of useful hourly events was 12654. Table 1 shows the number of hourly events used for each of the three periods of analysis, which were split, firstly into signal and background events, where signal events are those for which Rn activity is more than 200 Bqm^{-3}, and background is less than that value, and then each set was split once more, into training and testing sample to be used in MVA analysis. Table 1 also shows the number of events used, and split, in periods with air condition operation on (AC), line pattern area on fig. 2(a), and air conditioning off (noAC) gray on fig. 2(a).

Table 1. Summary table of number of hourly events used for specific part of analysis

	noAC	AC	All period
Signal training	1343	912	3428
Signal testing	1343	912	3428
Signal training and testing	2686	1824	6856
Background training	942	1531	2899
Background testing	942	1531	2899
Background training and testing	1884	3062	5798

Before performing the multivariate (MVA) analysis, we have looked into single variable analysis, and the best way to see if variables could be useful for analysis is if they have, firstly, the greatest correlation with radon activity (concentration), and, secondly, which variable profiles for

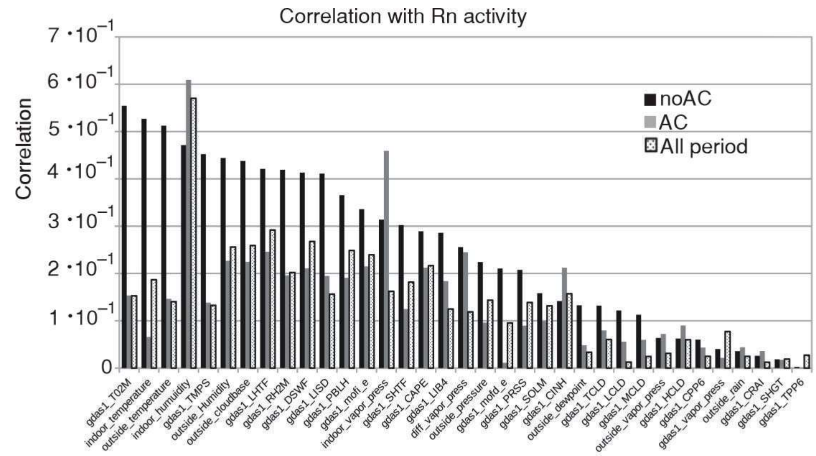


Figure 3. The modulus of Pearson's correlation coefficients of radon activity with each of variables used in the analysis is shown. Note the decreasing of correlation with temperature variables, and increasing with humidity variables, when air conditioning (AC) was turned on

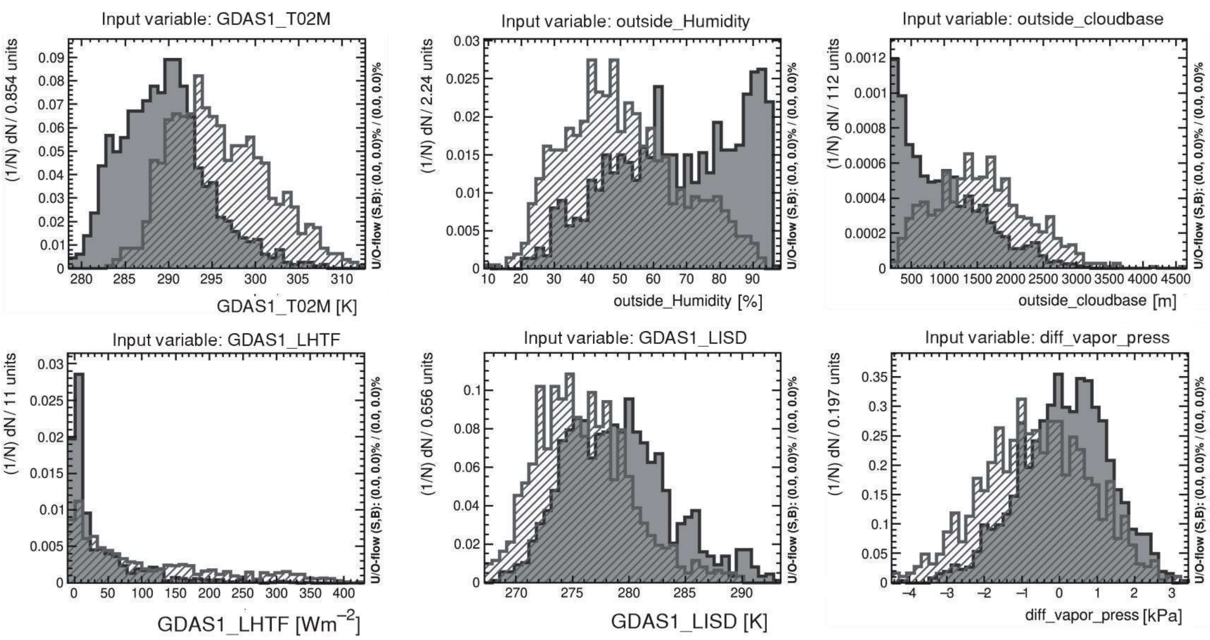


Figure 4. For some variables there is a significant separation of distributions of variables' values for events with low and events with high radon activity. Variables shown are: temperature at height of 2 m above the ground (GDAS1-TO2M), outside relative air humidity, measure of lowest visible part of the cloud (cloudbase), latent heat net flux at the surface (LHTF), standard lifted index (LISD) and the difference of water vapor pressure from indoor and outdoor

high Rn activity (signal) and low (background) data samples, have smallest overlap, meaning that they have greatest separation of high and low Rn activity samples. So, firstly, we are looking into modulus of Pearson's correlation coefficients for each of the variables used in this analysis with radon activity, fig. 3. Since the greatest variation of radon activity should give the best insight into correlation with variables, we are firstly looking into data with air condition off (noAC). To the variables with greatest modulus of Pearson's correlation coefficients with Rn activity (noAC) are temperature variables from all three sources of data GDAS, radonometar and meteorological station (GDAS1-T02M, indoor-temperature, outside-temperature, GDAS1-TMPS), than humidity (indoor-humidity, outside-humidity), outside-cloudbase, followed with GDAS variables: GDAS1-LHTF (latent heat net flux on surface) and GDAS1-DSWF (downward short wave radiation flux) and GDAS1-RH2M (relative humidity at height of 2 m), followed by indoor-vapor-pressure. When air conditioning is turned on, there is a change in correlation, where temperature variables correlations are decreasing, and there is an increase in correlation of humidity variables like indoor-humidity and indoor-vapor-pressure. We observe this change since temperature is now holding at approximately the same level by air conditioning, and any variation of radon activity we see does not come from approximately constant temperature. We noticed the similarity in modulus of Pearson's correlation coefficients of outside-T02M and outside-temperature with Rn activity of 55.4 % and 51.2 %, respectively, for noAC data, and 15.3 % and 14.6 %, respectively, for AC data. Also, outside-humidity and gdas1-RH2M with 44.4 % and 41.9 %, respectively, for noAC and 22.7 % and 19.6 % for AC data. When looking into pressure data, outside-pressure and GDAS1-PRSS have modulus of Pearson's correlation coefficients of 22.4 % and 20.8 %, respectively, for noAC data and 9.6 % and 9.0 % for AC data.

When looking into separation of variables for signal and background samples, fig. 4 shows selected

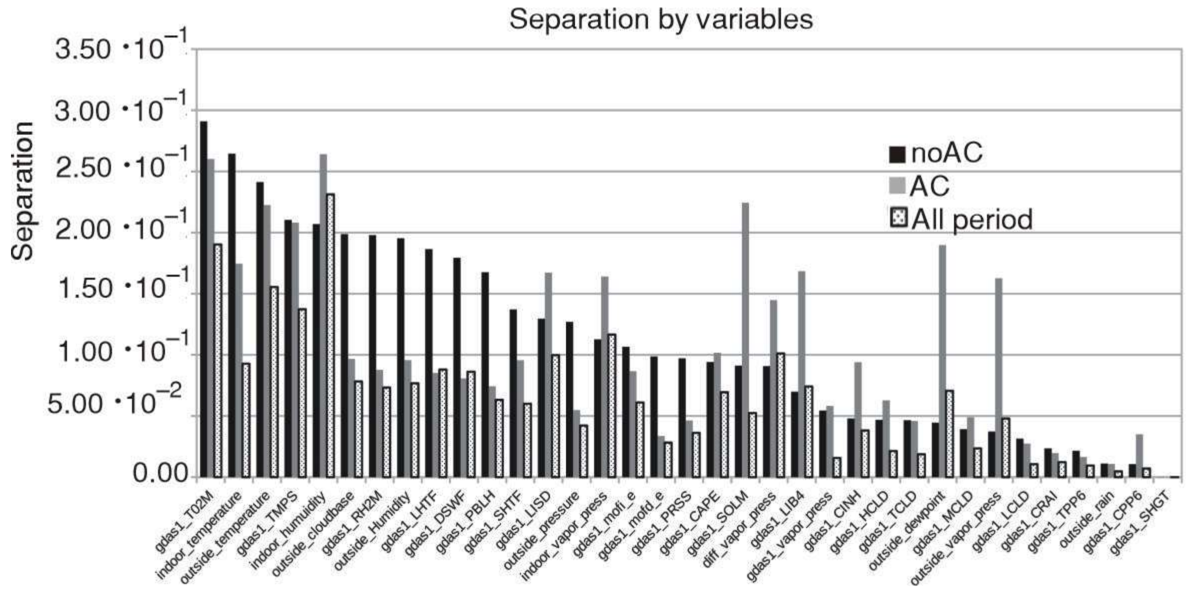


Figure 5. Separation of events with low and high Rn activity by each variable

variables, where separation can be seen with naked eye, and also, separations of high and low Rn activity for different variables can be roughly compared. But, we want to have more precise insight into separation, and for all three samples AC, noAC and samples of whole measurement period. This is shown in fig. 5 where we can see that for noAC, temperature variables have most significant separation values, as was the case with modulus of Pearson's correlation coefficients with Rn activity on fig. 3. With air conditioning turned on, the variables of humidity and vapor pressure gain in separation value, while indoor temperature is decreasing its separation value. Notice that the change is not so pronounced as was the case with correlation variables. Again, we noticed the similarity separation values of outside-T02M and outside-temperature 29.1 % and 24.1 %, respectively, for noAC data, and 26.0 % and 20.8 %, respectively, for AC data. Also, outside-humidity and GDAS1-RH2M with 19.8 % and 19.5 %, respectively, for noAC and 8.8 % and 9.6 % for AC data. When looking into pressure data, outside-pressure and GDAS1-PRSS have separation values of 12.7 % and 9.7 %, respectively, for noAC data and 5.5 % and 4.6 % for AC data.

MULTIVARIATE CLASSIFICATION ANALYSIS

Toolkit for multivariate analysis (TMVA) [11] implemented in ROOT [12] framework for data analysis, has many of multivariate methods and tools implemented, which are frequently used for data analysis, as in High energy physics, also by data scientists in general. We will not get into details of wide spread of multivariate methods available, which can be found in TMVA manual [11]. The usage of those multivariate methods in TMVA is rather standardized. What is advantageous in using TMVA is that we could compare many of multivariate methods using the same training and testing sample. Also, the TMVA was used in many analyses, and is constantly under development, with many new methods implemented. The TMVA offers comparison of methods developed for other frameworks, like methods developed in programming langnages Python, or R, or modern methods like Deep and Convolutional Neural Networks, which is best to be run in multi-thread mode or on CPU or on GPU (graphical cards).

In MVA analysis, the data sample consists of events. Event is composed of data measured/recorded at the same time for each input variable. We can run MVA as Classification, Classification with category, and Regression. The MVA Classification is done when sample is divided into two samples (classes); signal and background. The MVA methods are trained to make the same classification using events they have not seen before, and their performance in classification is measured. Second MVA analysis is done as regression analysis. It is similar to classification, in the sense that the number of classes into which initial sample is divided is much bigger, and the value of classifier is not only 1 (signal) and 0 (background) but has much more values in between. Classification with category was not used, as the maximum performance of Classification is obtained when no other categorical values besides 1 (signal) and 0 (background) are used. Future performance tests could include categories like; very high, high, medium, low and very low radon concentrations.

When a sample is prepared, MVA classification needs some time to complete the training process for

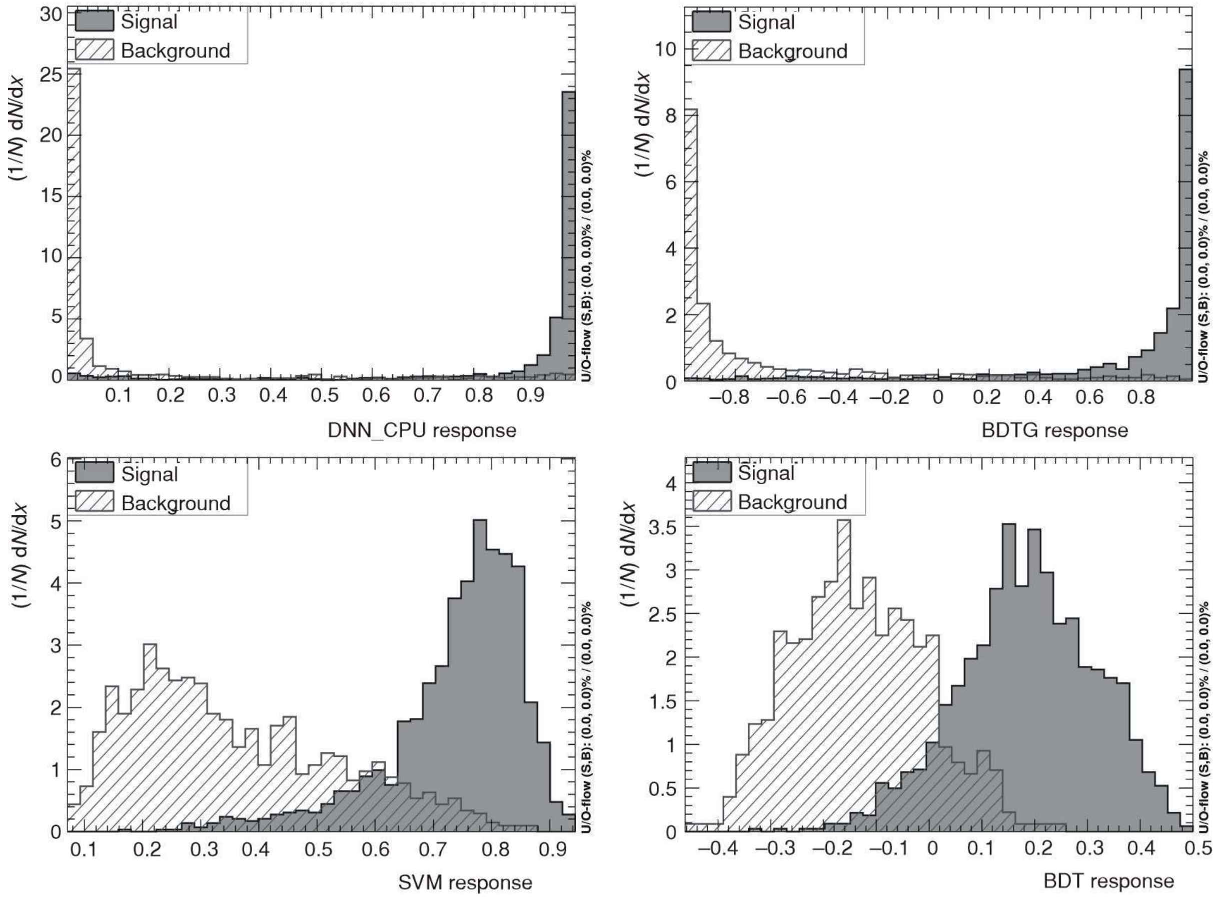


Figure 6. Response of MVA methods to events with low and high Rn activity

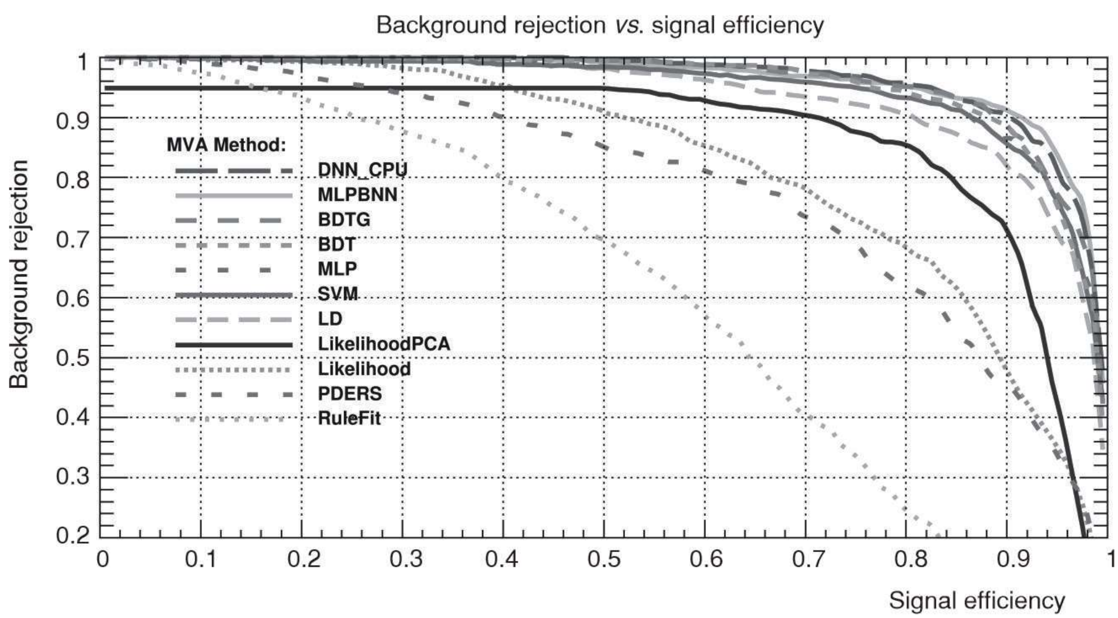
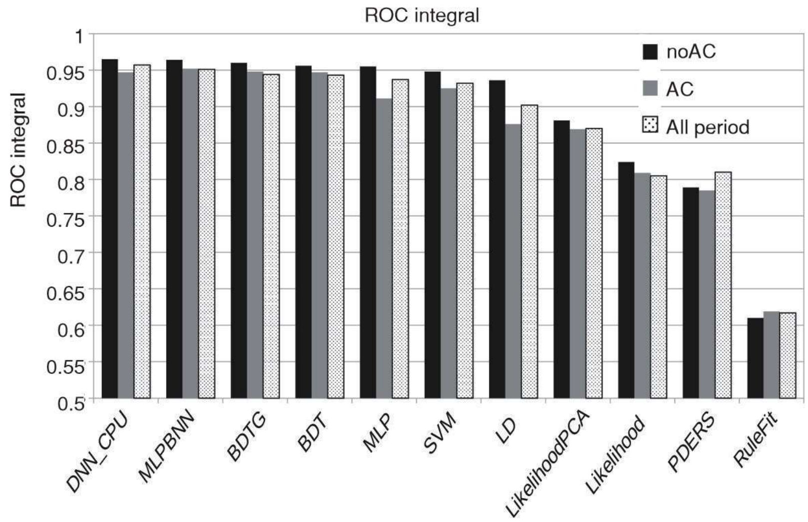


Figure 7. The ROC curve for MVA methods for the time interval where air conditioning was off (noAC)

each of MVA methods selected for comparison. Besides training, the sample of same number of events is used for evaluation, or testing, where MVA method is tested on samples not seen before (not used for training). The performance of some MVA method is expressed only using testing/evaluation sample.

The fig. 6 shows the response of best performing MVA methods, in analysis of noAC data, to events with low and high Rn activity, or signal and background. We can see, in fig. 7, that by looking into Receiver Operating Characteristic (ROC) curve comparison of all selected multivariate methods, that several

Figure 8. Value of the ROC integral for MVA methods for the selected time intervals, AC and noAC, and for whole time interval



methods have very good performances and also, very close performances. It is very good to have several methodologically very different multivariate methods performing in similar way, since this gives us confidence that classification is applicable. To illustrate this point, we can say that, very generally speaking, ANN are based on convolution of selected function to the resulting multivariate functional dependence, while Boosted Decision Trees are based on multidimensional space (cube) cuts, for approximation of multivariate functional dependence, and it is very good that both have very good performances in MVA classification.

The comparison of ROC curve integrals for best performing methods, for MVA classification analysis for all three intervals; noAC, AC and all-period analysis is shown at fig. 8. For five best performing methods, DNN-CPU (Deep Neural Network), MLPBNN (Multi-Layer Perceptron Bayesian regulator Neural Network), BDTG (Boosted Decision Trees with Gradient boosting), BDT (Boosted Decision Trees), and MLP (Multi-Layer Perceptron – an ANN), results are very similar, and also for all the three intervals, which is very important in sense that while variables' correlation with Rn activity vary greatly, this is easily overcome in MVA methods, adding very important property of robustness in variable selection. We should note that all the mentioned methods are ANN or DBT based multivariate methods.

The resulting trained multivariate methods are now ready to be included into some web applications, or used in variables' analysis. In web applications, Radon alarm could be constructed, when based on input variables, there is a great probability of increased indoor radon activity. For example, some places where it is known from previous measurements, like from participation in large indoor radon survey, that dwelling or working space has a problem with increased indoor radon concentration, some measures like increased ventilation or longer brakes from work, could be made. In variables' analysis, the simplification of MVA approximation of underlying multivariable function dependence could be made, not only with classification, but more effectively with regression methods.

The MVA methods which are trained and tested using full set of variables and all available data are ready to be used in some application. But, we can continue our work and try to modify something in our analysis chain to see if we can get better performance or method which uses lower number of input variables, without big loss in performance. We can make different selection of training data sets, like truncation of outlier data, we can change the number of input variables, or change parameters specific for each MVA method. For this purpose, it could be very useful to look into variable importance for specific MVA method, for example for BDTG in fig. 9, in order to look into the influence of variables on MVA decision. To show why this is useful we pay attention on Pearson's correlation coefficients of input variables and radon concentrations and notice that there could be several variables with high correlation coefficient with radon concentration, but highly inter-correlated with each other, which results in no gain in MVA method performance if we add several variables which are inter-correlated. So, we can exclude variables if their exclusion does not lower the MVA method performance. We choose to look into importance of variables on BDTG classification, for all time intervals. Again, we start with noAC intervals, where indoor radon activity was highest, and indoor temperature was not regulated. We start with two GDAS variables, GDAS1-SHTF (sensible heat net flux at surface) and GDAS1-SOLM (volumetric soil moisture content), followed by indoor-humidity and diff-vapor-pressure, and GDAS1-T02M at position 6, with some other variables similarly important as gdas momentum flux direction and gdas cloud cover variables.

When comparing data from meteorological station and gdas data, we cannot compare them in, for example, multivariate importance, since if one variable is chosen to be used in MVA training, similar variable in, for example Pearson's correlation coefficients or

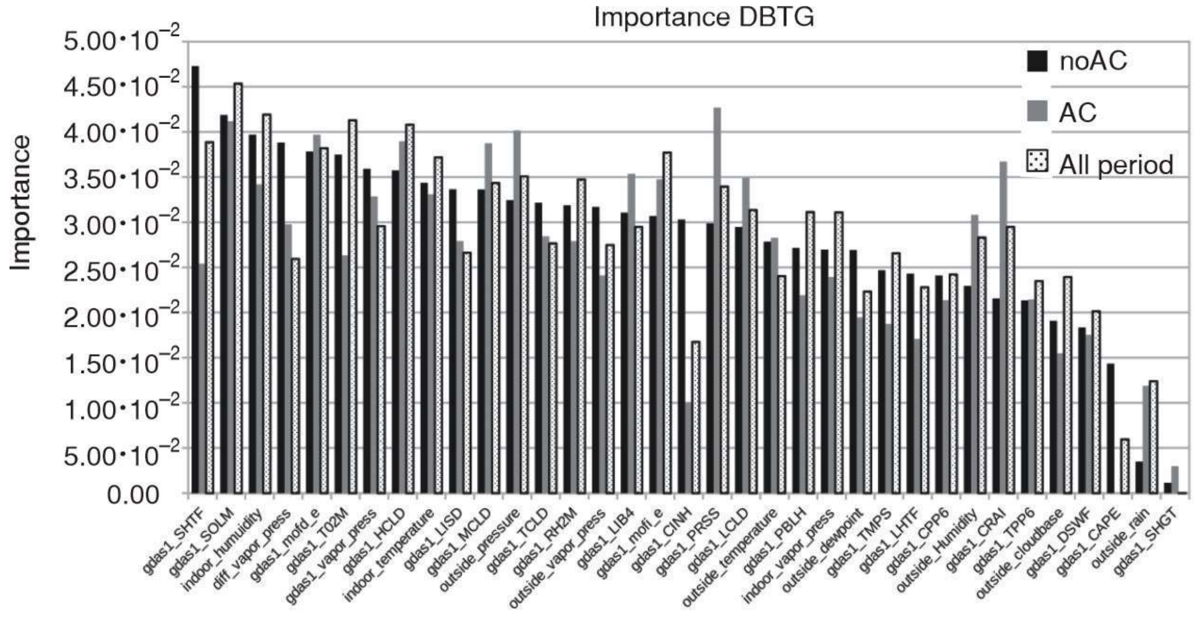


Figure 9. Variable importance for MVA method BDTG for time intervals, AC and noAC, and for the whole time interval

separation of variable for increased and for low Rn activity value, do not have power to make discrimination. Comparison can only be used when each variable is observed separately in a single variable analysis. Also, similar situation can happen with preparation of variables, where resulting variables are, de-correlated, and first variable is significant for further analysis but other, very similar variable before de-correlation, remains with negligible significance for further analysis.

THE MVA REGRESSION

Regression analysis often fails if there is not a strong dependence of target variable, in our case Rn activity, on input variables. Reasoning is the following: Classification analysis has only two outputs, either it is signal (1) or background (0), but in case of regression, there are many more values between 0 and 1, and much more dependence, or events is needed to get positive results here. We ran MVA regression for three time intervals, noAC, AC and all-period. The BDTG and DNN-CPU show good prediction results after MVA regression training procedure, as a result of RMS of deviations of true and evaluated value of Rn activity are satisfyingly small, as is shown in fig. 10. The fig. 11 shows this in more detail for BDTG in noAC regression analysis, where the distribution of deviations is shown for each event in the testing sample.

CONCLUSIONS

Single variate analysis of correlations of each of meteorological variable with indoor radon activity and Multivariate classification and regression analysis of all meteorological variables and radon activity was per-

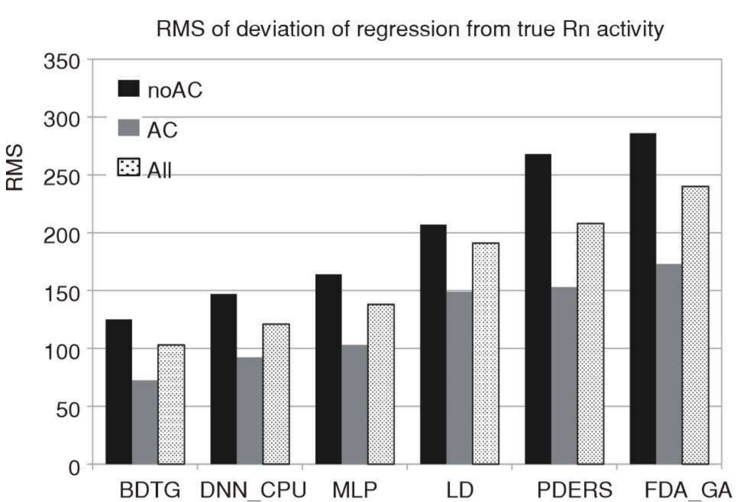


Figure 10. The RMS of deviations of regressions from true value for selected time intervals, AC and noAC, and for the whole time interval, for several MVA regression methods

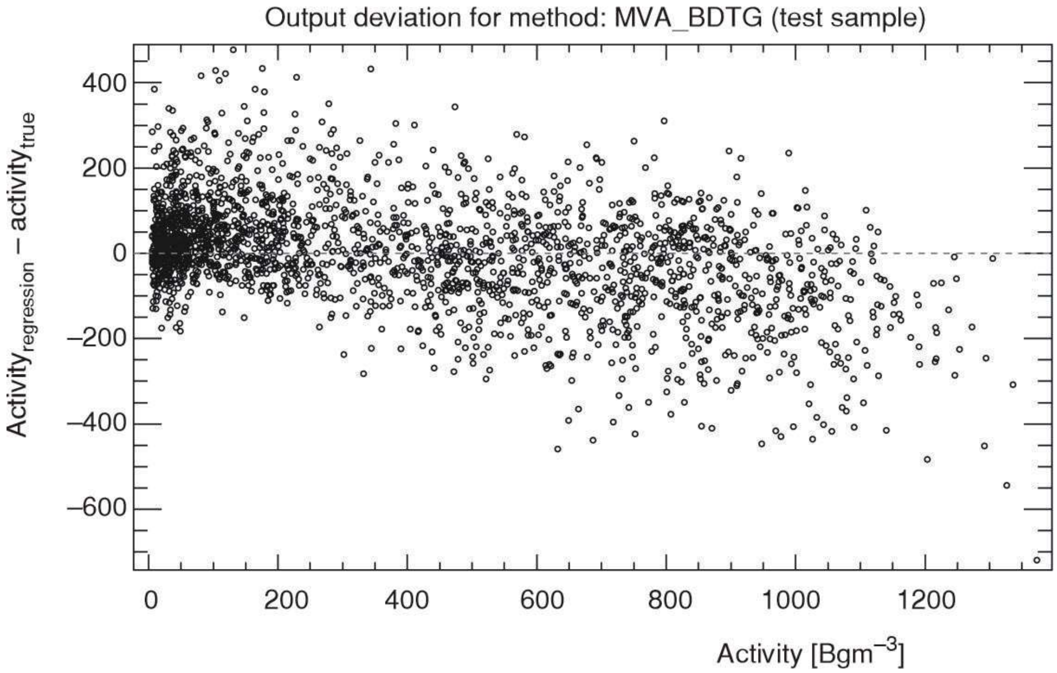


Figure 11. Deviation of regression from true value for noAC period and BDTG MVA method

formed and discussed. Meteorological variables used in this analysis were from radonometar device, then from a nearby meteorological station and finally from GDAS data. Single variate analysis has identified variables with greatest value of modulus of Pearson's correlation coefficient with Rn activity, and also variables with greatest separation of events with increased Rn activity of over 200 Bqm^{-3} and of events with Rn activity below this value. This initial analysis and looking into variables were showing the expected behavior of Rn concentration with meteorological variables, with emphasis on data periods with or without air conditioning, and also with emphasis on previously found variable of indoor water vapor pressure. This single variate analysis and observing all the data proved also useful for conclusion that GDAS data could be used as a good enough approximate replacement for meteorological data from the nearby meteorological station for MVA analysis. The MVA classification analysis found several very well performing MVA methods which can be used in web application or for further detailed analysis of specific input variables. Variable importance of BDTG MVA method was shown for all three periods, and most important variables were discussed. Finally, MVA regression analysis gave also good results, and more quality measurements in this rarely accessed ground level laboratory would be useful to better tune the MVA methods, and do more detailed analysis.

ACKNOWLEDGMENT

Authors thank the NOAA Air Resources Laboratory (https://www.arl.noaa.gov/) for GDAS data. (https://www.ready.noaa.gov/gdas1.php). Authors thank the Environmental Physics Laboratory (http://www.envpl.ipb.ac.rs/) for meteorological station data. The authors acknowledge funding provided by the Institute of Physics Belgrade, through grant by Ministry of Education, Science and Technological Development of the Republic of Serbia.

AUTHORS' CONTRIBUTIONS

The original idea and draft were carried out by D. M. Maletić. The data provided by R. M. Banjanac, V. I. Udovičić and Z. Mijić. Statistical analysis was done by D. M. Maletić, D. R. Joković and A. L. Dragić. N. B. Veselinović, M. R. Savić, S. Živković-Radeta and J. V. Udovičić worked on data preparation and selection. All the authors analyzed and discussed the results and reviewed the manuscript.

REFERENCES

[1] ***, WHO Handbook on Indoor Radon: A Public Health Perspective, World Health Organization, Switzerland, 94, 2009
[2] Dragić, A., *et al.*, The New Set-Up in the Belgrade Low-Level and Cosmic-Ray Laboratory, *Nucl Technol Radiat, 26* (2011), 3, pp. 181-192
[3] Maletić, D., *et al.*, Comparison of Multivariate Classification and Regression Methods for the Indoor Radon Measurements, *Nucl Technol Radiat, 29* (2014), 1, pp. 17-23
[4] Eremić-Savković, M., *et al.*, Results of the First National Indoor Radon Survey Performed in Serbia, *Journal of Radiological Protection, 40* (2020), 2, pp. N22-N30

[5] Vukotić, P., *et al*., Influence of Climate, Building, and Residential Factors on Radon Levels in Ground-Floor Dwellings in Montenegro, *Nucl Technol Radiat, 36* (2021), 1, pp. 74-84

[6] Udovičić, V., *et al*., Multiyear Indoor Radon Variability in a Family House – a Case Study in Serbia, *Nucl Technol Radiat, 33* (2018), 2, pp. 174-179

[7] Udovičić, V., *et al*., Radon Variability Due to Floor Level in Two Typical Residential Buildings in Serbia, *Nukleonika, 65* (2020), 2, pp. 121-125

[8] Maletić, D., *et al*., Correlative and Multivariate Analysis of Increased Radon Concentration in Underground Laboratory, *Radiation Protection Dosimetry, 162* (2014), 1-2, pp. 148-151

[9] Tetens, O., About Some Meteorological Terms (original: Über Einige Meteorologische Begriffe.) *Z. Geophys, 6* (1930) pp. 297-309

[10] Murray, W., On the Computation of Saturation Vapor Pressure, *J. Applied Meteorology, 6* (1967), pp. 203-204

[11] Hoecker, A., *et al*, TMVA – Toolkit for Multivariate Data Analysis. *PoS ACAT, 40* (2009), p. 12

[12] Brun, R., Rademakers, F., ROOT – An Object Oriented Data Analysis Framework, *Nucl. Inst. & Meth. in Phys. Res. A, 389* (1997), pp. 81-86

Received on December 18, 2023
Accepted on February 12, 2024

Димитрије М. МАЛЕТИЋ, Радомир М. БАЊАНАЦ, Дејан Р. ЈОКОВИЋ, Александар Л. ДРАГИЋ, Никола Б. ВЕСЕЛИНОВИЋ, Михаило Р. САВИЋ, Зоран Р. МИЈИЋ, Владимир И. УДОВИЧИЋ Светлана Д. ЖИВКОВИЋ-РАДЕТА, Јелена В. УДОВИЧИЋ

МУЛТИВАРИЈАНТНА АНАЛИЗА ДВОГОДИШЊЕГ КОНТИНУАЛНОГ МОНИТОРИНГА РАДОНА У НАДЗЕМНОЈ ЛАБОРАТОРИЈИ У ИНСТИТУТУ ЗА ФИЗИКУ У БЕОГРАДУ

Приказана је мултиваријантна класификациона и регресиона анализа односа метеоролошких варијабли и концентрације радона у затвореној и ретко приступачној приземној лабораторији Института за физику Београд. Податке о метеоролошким варијаблама и концентрацији радона, коришћене у овој анализи, добијамо из активног уређаја за краткорочна мерења концентрације радона у затвореном простору, оближње метеоролошке станице и из података Глобалног система асимилације података. Једно-варијантном анализом идентификоване су варијабле са највећом вредношћу модула Пирсоновог коефицијента корелације са концентрацијом радона, као и варијабле са највећом моћи раздвајања догађаја са повећаном концентрацијом радона више од (200 Bqm^{-3}) и догађаја са нижом концентрацијом од ове вредности. Ова почетна анализа и сагледавање варијабли показују очекивану везу концентрације радона и метеоролошких варијабли, са нагласком на анализу података из различитих временских интервала, када је у лабораторији радила и када није радила климатизација, као и са нагласком на варијаблу разлика унутрашњег и спољњег притиска водене паре. Ова једно-варијантна анализа доводи до закључка да се подаци Глобалног система асимилације података могу користити као довољно добра приближна замена за метеоролошке податке из оближње метеоролошке станице за мултиваријантну анализу. Мултиваријантном класификационом анализом пронађено је неколико веома добрих мултиваријантних метода које се могу користити у некој веб апликацији или за даљу детаљну анализу специфичних улазних варијабли. Приказана је важност варијабли за мултиваријантни метод стабла одлучивања за сва три периода мерења, а разматране су и најважније варијабле. Коначно, мултиваријантна регресиона анализа је такође дала добре резултате, што може да буде корисно при оптимизацији класификационих мултиваријантних метода.

Кључне речи: континуирани радон мониторинг, мултиваријантна анализа, Глобални систем асимилације података, метеоролошка станица