Audit financiar, XVI, Nr. 3(151)/**2018**, *382-397*
ISSN: 1583-5812; ISSN on-line: 1844-8801

An interdisciplinary analysis with data mining and visualization tools applied on multiple and multi-source time series

– The case of the forest fund in Romania –

Daniel HOMOCIANU,
E-mail: dan.homocianu@gmail.com
Dinu AIRINEI,
E-mail: adinu@uaic.ro
Ciprian Ionel TURTUREAN,
E-mail: ciprian.turturean@uaic.ro
Alexandru Ioan Cuza University of Iasi, Faculty of Economics and Business Administration, The Research Department and The Department of Accounting, Business Information Systems and Statistics,

Abstract

Starting from the idea of poor biodiversity conservation which continues to be one of the major both national and international current problems, this paper focuses on the analysis of the evolution of the forest fund and also on the validation of a set of hypothesis regarding specific indicators related to this evolution with a special consideration for interrelations and future trends based on a custom data mining model. The time series data was provided by the Romanian Ministry of the Environment, National Bank of Romania, Romanian Court of Accounts, National Institute of Statistics, Eurostat, Global Forest Watch, Global Forest Change, Google Earth and Transparency International. This data was filtered, calibrated, exported in order to be collected and integrated into a single online spreadsheet shared for open access and then progressively refined. It served as source of online representations under the form of interactive charts and maps or other visual metaphors able to capture the space-time evolution and distribution and provide basic support for querying. Finally, it was also the starting point for further time series analysis by using data mining tools. The results confirm the hypotheses formulated also on the basis of worrying satellite data and supported by the scientific literature. They also emphasize the influence of some historical moments associated with the appearance of specific laws progressively allowing an increasing private share of forests and also with the Romania's alignment to the European Union on the evolution of some forest fund indicators.

Keywords: *biodiversity conservation, forestry, interdisciplinary approach, interactive visualizations, multiple time series, data mining.*

JEL Classification: *C32, C88, Q23, Q57, Y10, Y91*

To cite this article:
Homocianu, D., Airinei, D., Turturean, C.I. (2018), An interdisciplinary analysis with data mining and visualization tools applied on multiple and multi-source time series – The case of the forest fund in Romania, *Audit Financiar*, vol. XVI, no. 3(151)/2018, pp. 382-397,
DOI: 10.20869/AUDITF/2018/151/018

To link this article:
http://dx.doi.org/10.20869/AUDITF/2018/151/018
Received: 25.05.2018
Revised: 14.06.2018
Accepted: 21.06.2018

Introduction

The economic sustainability (Business Dictionary, 2018) involves the use of various strategies for employing existing resources optimally so that a responsible and beneficial longer-term balance can be achieved.

In a business context, the economic sustainability translates into the efficient use of company's assets to enable it to continue to profitably operate over time.

Moreover, when speaking of sustainable economy, some authors (Chouinard et al, 2011) consider that even those who are only worried about business and not the fate of the planet, admit that the viability of the business itself depends on the resources of healthy ecosystems (fresh water, clean air, robust biodiversity, productive land) and the stability of fair societies.

When referring to the circular economy in the context of the shifting economy, other authors (Mossay and De Kemmeter, 2017) state that most inventory costs should disappear in the context of the first one because on can convert trash into resources for new products. They also analyze this type of circular economy (recycling / zero waste) together with some other models referred as next economies and based on characteristics such as: solidarity for the common good (social economy), open source and creative commons (knowledge economy), big data, internet of things and artificial intelligence based (digital economy), collaborative / cooperative / peer-to-peer (sharing economy), performance (functional economy), symbiosis (systemic economy), fair trade (ethical economy), restoration / regeneration (green economy) and biodiversity / agroforestry / healthy food production oriented (biological or simply bio-economy).

The General Directorate for Research and Innovation of the European Commission (EU Publications, 2015) considers that the bio-economy refers to the sustainable production of renewable biological resources and waste streams conversion into food, feed, bio-based products such as bio plastics, biofuels and bioenergy. Thus it includes agriculture, forestry, fisheries, food, pulp and paper production, and parts of chemical, biotechnological and energy industries.

As suggested in its title, this paper has as main goal the analysis of the evolution of the Romanian forest fund. For that it uses visual representation and data mining tools and focuses on a set of variables (see Fig.8 and the list of acronyms and abbreviations) considered to be closely related to this evolution and tries to validate a set of hypotheses. Consequently it uses notions, results and methods from many disciplines such as: Forestry, Economics, Accounting and Audit, Law, Geographic Information Systems, Statistics and Business Intelligence.

Nowadays the effective data visualization is considered (Naish and Burns, 2017) to be the bridge between the quantitative content of the data and the human intuition and has as purpose the insight (Card et al, 1999), not the pictures themselves. Therefore the necessity to provide support for user interaction and to present information in various ways including different levels of synthesizing / details becomes self-evident.

In order to formulate the hypotheses of this study we started by analyzing some simple ideas in this area. Later in the paper we present some interactive and online self-explanatory representations serving as basic support for emphasizing some inter-correlations and the evolution in time (scatter charts / line charts). The interactivity so strongly encouraged in recent years by the major publishing houses (Elsevier, 2018) is assured in the aforementioned online representations mainly by three means: (1) check-boxes for series; (2) the x-axis with a corresponding slider filter for both left and right edges; (3) direct feedback as callouts with data details when the pointing device is over a certain point on the chart surface.

The recent evolution of forest loss in Romania is full of evidence of illegal cuts including the reports from the Greenpeace organization which indicate from 30 to 96 cases per day (Green Peace, 2017) between 2009 and 2015. Moreover, this organization published information according to which (Green Peace, 2015) Romania lost 3 hectares of forest every hour in 2015. Therefore our first hypothesis (H1) admits a value close to the aforementioned one, but considers it to be an average for at least 15 years and having a strong relation with at least one type of forest cutting officially reported.

If considering the data about the forest fund from a report since 2013 belonging to an independent Romanian institution having the role of controlling the use of financial resources of the state, public sector and European community (Court of Accounts, 2013), we must accept an evolution including older time references, namely 1800:8500 K ha (thousands hectares), 1920:7880 K ha, 1929:6679 K ha, 1948:6486 K ha. Therefore, we have evidences (Google Sheets, 2018a) of a general descending trend except the approximately last forty years (last 6 time references: 1974:6332 K ha, 1985:6343 K ha, 1990: 6367 K ha, 2009:6496 K ha, 2010:6515 K ha, 2011:6519 K ha) which indicate vice versa. There are even studies (Gabryjończyk and Kułaga, 2017) proving a statistical relationship measured by correlation, between the size of forests, protected areas and surface of inland waters, as well as their share in the overall area of administrative units, and the number of beds in tourist accommodation establishments and their occupancy rate. There are also studies (Fodor and Pică, 2012) which emphasize the legislation's impact on the status of protected areas mainly due to need for harmonization with the ones of wider communities. In this context, our second hypothesis (H2) considers that the recent ascendant evolution of the Romania's protected areas strongly depends on the moments which preceded the Romanian join (2006-2007) in European Union (EU) in relation with the total area with forests and the number of tourist accommodation units, all having an overall ascending trend. Moreover both the protected areas and the total area with forests suddenly changed their previous growth (GAS, 2018a) and dependence (GAS, 2018b) pattern just before the aforementioned moments.

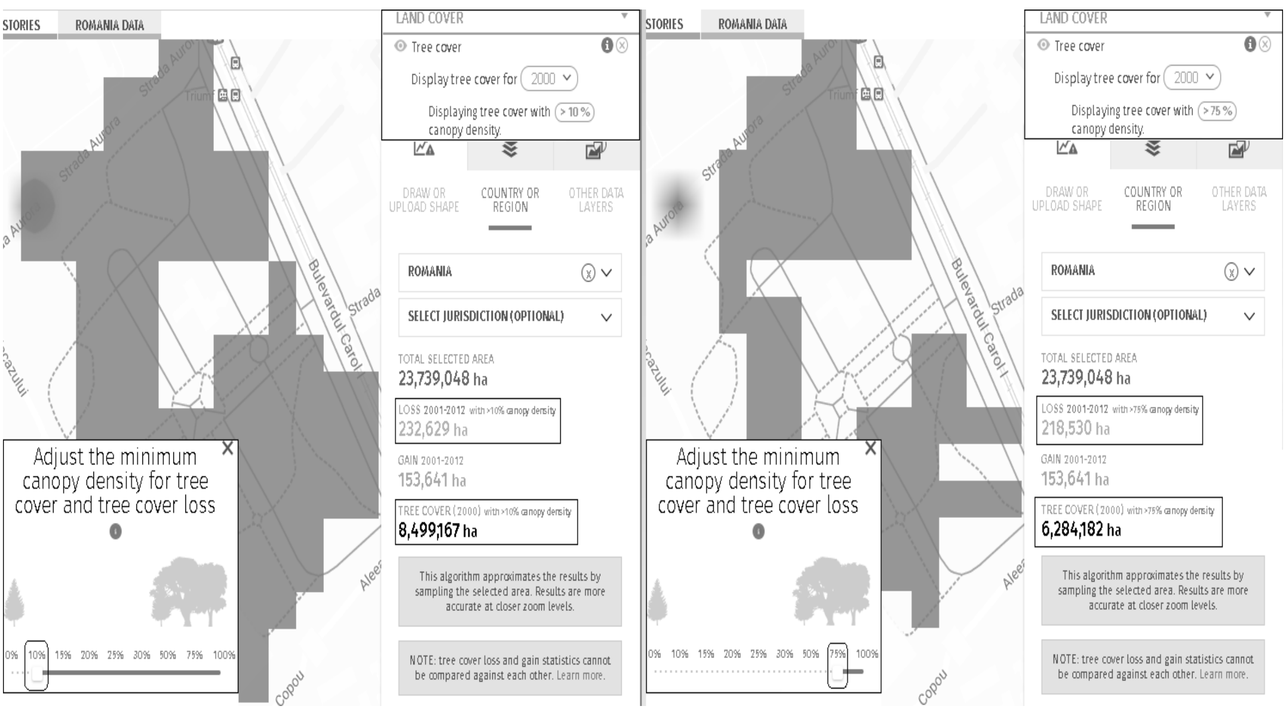
The principles of regeneration sylviculture (Johnson et al, 2009) state that the names of regeneration methods are also commonly used to refer to the type of harvest cutting involved. Moreover, according to another point of view (Nichiforel, 2014), the amount of wood resulting from cutting and harvesting mature trees form the group of main forest products and the corresponding cuttings are called main product cuts or regeneration cutting. Therefore our third hypothesis (H3) starts from the strong relation (GAS, 2018c) between the evolutions (GAS, 2018d) of the overall volume of harvested wood and that of the areas covered by regeneration cutting. We assume that both begun to change their evolution patterns after a certain time landmark belonging to the period between the adoption of two important laws concerning the restoration of the property rights on agricultural and forest land, namely Law No.1/2000 and the controversial Law No.400/2002. According with the same report of the Court of Accounts, the first one started to generate effects a year later causing both a sudden and a long term increase of the private property share on Romanian forests due to the raise of the individual quota and that of the number of allowed archaic associative forms of ownership – also proven by the great slopes of the lines that link 2000 with 2001 and 2000 with 2002 (GAS, 2018e). The second one was adopted in the same year (2002) an Austrian wood processing giant (Holzindustrie Schweighofer, 2014) entered the Romanian market and the National Anti-corruption Prosecutor's Office (PNA) was established. The last one was later known as the National Anti-corruption Department and finally Division (DNA) and was a product of an Emergency Ordinance (No.43/2002) of the Romanian Government succeeding an important law, namely Law No.78/2000 on the prevention, detection and sanctioning of corruption.

According to the conclusions section of an official report published by the National Institute of Statistics from Romania / INSSE (NIS, 2011) and containing data about the energy consumption in 2009 expressed in TJ (Terra Joules), the firewood (including biomass) used in households had a corresponding share of 49.64% from all types of fuels while in 1996 the same share was lower (43.08%). That difference was partially explained also by different decrease patterns of Romania's population and number of occupied dwellings. In terms of history of industrial civilization as history of energy transitions, according to an well documented point of view (Timmons et al, 2014), as supplies of firewood and other biomass energy proved insufficient to support growing economies, people turned to coal and then to oil, natural gas, hydro, nuclear, wind, direct solar and geothermal power. If considering civilization in general, there are well-founded views (Bailey, 2007) stating that: (1) cultures, civilizations and communities develop preferred horizons similarly to the way individuals do; (2) the governments' laws and policies are in fact a reflection of the spiritual maturity of the society they belong to; (3) different societies have different levels of spiritual consciousness depending on their inhabitants' preferences; (4) any civilization rises and falls in step with its changing collective consciousness and when corruption in any community becomes systemic, individuals within that society are more at risk and the civilization suffers. Moreover, economists (Boone and Kurtz, 2014) refer to the nations' employment / unemployment rate as an indicator of their economic health. The last one together with the personal and national prosperity and security (including physical), both subject of the civilization paradox (Dal Bó et al, 2015), are profoundly affected by the geography (Chapman, 2011). From a certain point of view closer to both the economic and law theory (Solum, 2013), the transparent markets together with the equality, prosperity and liberty as preconditions for human flourishing are at risk and cannot long persist in a society with a corrupt judiciary. In this context, our fourth hypothesis (H4) considers that the evolution of the volume of internally produced firewood / biomass as a primary energy resource usually known as not enough (Ministry of Waters and Forests, 2017) for what the population consumes is strongly related with that of the corruption perception and the overall employment rate (Skaf, 2011) of labor resources in Romania, all three acting as basic measures of the level of civilization. Moreover we assume an inverse relation between the amounts of turnover of the forestry units and the employment rates.

1. Data and methods

We have used tabular data from many sources such as: Eurostat, NIS, and the Transparency International (TSPI) - Global Coalition against Corruption. We have also used results from time-series analysis of Landsat images about tree cover (TC) and tree cover loss (TCL) provided by the University of Maryland (Hansen et al, 2013) / Google / USGS (United States Geological Survey) / NASA (National Aeronautics and Space Administration) and accessed through Global Forest Watch (GFW) and Global Forest Change (GFC). The Landsat (Google Earth Engine, 2018) is actually a joint program of USGS and NASA which has been observing the Earth continuously from 1972. Today the Landsat satellites provide imagery data for the entire Earth's surface at a 30x30 meter resolution about once every two weeks, including multispectral and thermal data. Earth Engine makes this data available in its raw form, as Top of Atmosphere (TOA) - corrected reflectance, and in various “ready-to-use” computed products such as Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI). Next we have compared the data provided by Eurostat and NIS with data from independent organizations with corresponding projects such as GFW and GFC. In the case of GFW we found a configuration parameter, namely the Minimum Canopy Density (MCD), to properly get and use data about TC and TCL.

Figure 1. Imagery data with TC and TCL results (2001-2012) depending on MCD



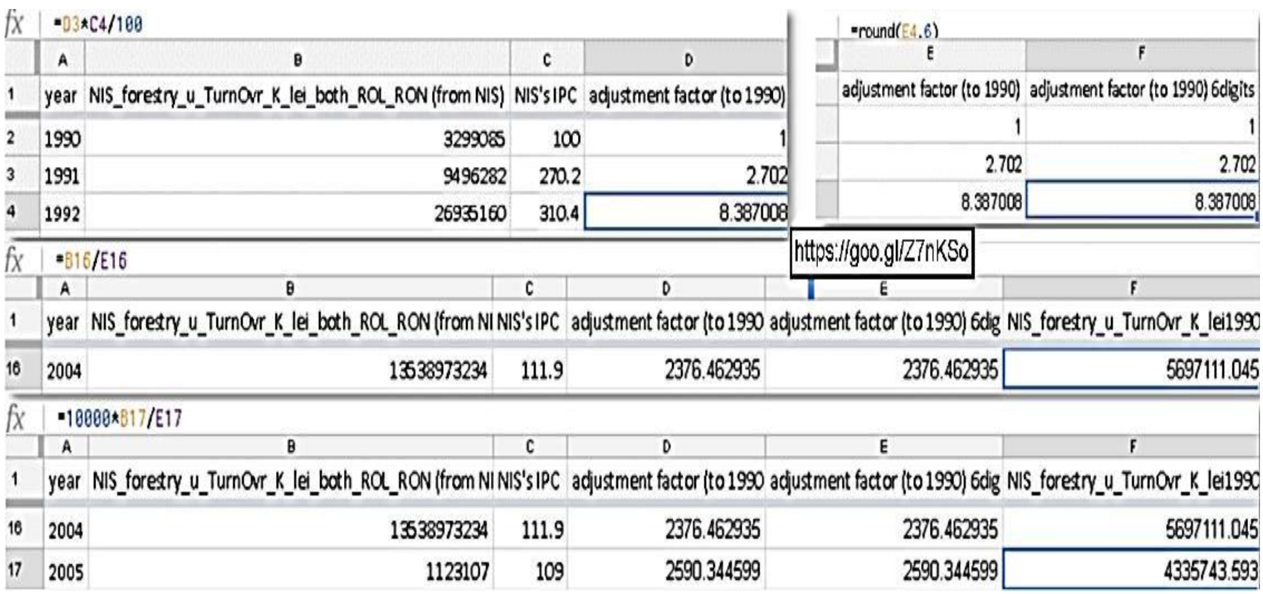
Source: GFW, 2018a

In fact, we have started from NIS and Eurostat reports for Romania in 2000 because this year was the only one available for calibration in GFW. Both NIS and Eurostat indicated a total forest fund of almost 6366000 ha (NIS, 2007), while the GFW's default report (MCD=30%) pointed out 7915160 ha (GFW, 2018b). Using the same sources and specifications, we did a calibration allowed by GFW when adjusting the MCD (Fig.1 - only seven thresholds: 10%, 15%, 20%, 25%, 30%, 50% and 75%) to a value able to cause a TC value for 2000 (6284182 ha) closer to the one reported by both NIS and Eurostat (minimum absolute value of the difference). The optimum MCD resulting from this single point data calibration (year 2000) was 75% which we have used as reference to get the correct TCL data presented below. In fact 75% was the only MCD threshold allowing a reasonable difference between NIS / Eurostat and GFW data (e.g. 81818 ha = 6366000-6284182 in favor of NIS or 1.3% of the total area with forests -TFA as reported by NIS in 2000). The rest of thresholds (first 6) caused huge differences for 2000 (between 1 and 2 million ha or >15.7% and <31.5% of TFA as reported by NIS in 2000) but in favor of GFW (opposite sign) which comes with stable engine specifications (Earth Engine Partners, 2013a) mentioning only trees taller than 5 meters, in accordance with the Food and Agriculture Organization of the United Nations (FAO, 2015)'s forest definition.

The resulting data set (Google Sheets, 2018b) has many additional columns and rows (1990-2016) finally serving as input for the Microsoft's Excel Data Mining (DM) add-in used with Microsoft SQL Server Analysis Services (SSAS) 2012, responsible for ensuring the model's persistence. The columns with data available for different periods of time included: the total forest area, expressed in K ha (NIS_tot_forest_area_K_ha or TFA), the forest area where artificial regeneration took place in K ha (NIS_artif_regen_forest_K_ha or ARF), the forest area where regeneration cutting occurred in K ha (NIS_cutting_regen_forest_K_ha or CRF), the forest area where hygiene and cleaning cutting occurred in K ha (NIS_hygiene_cleaning_cutting_K_ha or HCC), the forest area where cuts of accidental products such as torn trees occurred in K ha (NIS_accid_prod_cutting_K_ha or APC), the employment rate of the labor resources in percentages (NIS_emplym_rate_of_lab_res or ERLR), the total number of tourist accommodation units / establishments (NIS_tourist_accommod_units or TAU), the amount of turnover for forestry units expressed in thousands of lei at their value in 1990 (NIS_forestry_u_TurnOvr_K_lei1990 or TO), computed using a processing / adjusting algorithm described below (**Figure 2**), the overall amount of harvested wood expressed in thousands of cubic meters (NIS_overall_harvested_wood_K_m3 or OHW), the overall amount of firewood including biomass from self-production expressed in thousands of tons (NIS_intern_firewood_biomass_K_to or IFB), the overall area of protected areas in K ha (NIS_protected_areas_K_ha or PA), the overall amount of water resources expressed in millions of cubic meters (NIS_total_water_resources_KK_m3 or TWR), the overall amount of wood sold by the forestry units in K of cubic meters (NIS_forestry_u_sold_wood_K_m3 or SW), the corruption perception index score on a 0-10 scale considering the higher the better (TSPI_corrupt_percept_idx_score or CPI), the overall corruption perception rank – the higher the worse for all countries including those from the EU (TSPI_CP_overall_rank or CPOR) and finally the TCL for vegetation greater than five meters in height and MCD >75% expressed also in K ha (GFW_trees_over5m_cover_loss_K_ha or TCL).

When using the Excel DM add-in we have included most of the columns available (the full dataset except last four columns about property share) and the default settings indicating a forecast method (Microsoft, 2017) using a mix of ARIMA (Auto-Regressive Integrated Moving Average) described in the literature of the field (Box et al, 2015) and ARTXP (Auto Regressive Trees with Cross Predict), latter as a method derived from ART (Auto Regressive Trees), also consecrated (Meek et al, 2002).

Figure 2. Formulas and functions used to generate comparable values of turnover



fx =D3*C4/100

	A	B	C	D
1	year	NIS_forestry_u_TurnOvr_K_lei_both_ROL_RON (from NIS)	NIS's IPC	adjustment factor (to 1990)
2	1990	3299065	100	1
3	1991	9496282	270.2	2.702
4	1992	26935160	310.4	8.387008

=round(E4,6)

E	F
adjustment factor (to 1990)	adjustment factor (to 1990) 6digits
1	1
2.702	2.702
8.387008	8.387008

https://goo.gl/Z7nKSo

fx =B16/E16

	A	B	C	D	E	F
1	year	NIS_forestry_u_TurnOvr_K_lei_both_ROL_RON (from NI	NIS's IPC	adjustment factor (to 1990	adjustment factor (to 1990) 6dig	NIS_forestry_u_TurnOvr_K_lei1990
16	2004	13538973234	111.9	2376.462935	2376.462935	5697111.045

fx =10000*B17/E17

	A	B	C	D	E	F
1	year	NIS_forestry_u_TurnOvr_K_lei_both_ROL_RON (from NI	NIS's IPC	adjustment factor (to 1990	adjustment factor (to 1990) 6dig	NIS_forestry_u_TurnOvr_K_lei199C
16	2004	13538973234	111.9	2376.462935	2376.462935	5697111.045
17	2005	1123107	109	2590.344599	2590.344599	4335743.593

Source: authors' work in an online spreadsheet, 2018a

The processing algorithm (**Figure 2**) for adjusting the data corresponding to the turnover for forestry units started from NIS data about the Annual Consumer Price Index (ACPI / IPC) based on comparisons of evolution of prices of purchased goods and tariffs of services used by the population in the current year compared to the previous one. The reference point was set for 1990 (a value of 100) and the rest of values were taken directly from a source table (NIS, 2017) with IPC data. Next steps were to compute the adjustment factors (to 1990) for each year (1991-2016) as a cascade product of IPCs divided by 100 and finally do divide the original amount of each turnover between 1991 and 2016 to the corresponding adjustment factor (**Figure 2**).

For last twelve time references (2005-2016) the resulting values were multiplied by 10000, meaning the multiplication factor between the old and the new monetary units in line with the monetary reform of 1 July 2005 (BNR, 2014), ended with the denomination of the Romanian currency. In fact, NIS's online app reported both ROL - Romanian Leu (1990-2004) and RON - Romanian New Leu (2005-2016) as generic Lei in a single data set export about forestry units' turnover (1990-2016) with no automatic adjustment able to take into account that one RON equals 10000 ROL.

2. Results and discussion

GFW comes with online visualization facilities similar in terms of dynamics to the offline ones provided by the Microsoft Excel Power Map add-in (Homocianu and Airinei, 2014a) component which in turn requires stand-alone dedicated graphical processing power. They are able to present the evolution both in space / geographical and time / years (**Figure 3**) by considering TCL vs. tree cover gain (TCG).

This interactive type of representations (Fig.3) is also the starting point when trying to validate / reject H1. Thus, for hourly loss corresponding to all days in a year we get 1.98 ha (277652 / 16 years / 365 days / 24 hours a day) for MCD=75%. For hourly loss corresponding to working days (Google Sheets, 2018c) in a year, we get 5.68 ha = 277652 / 16 years / 254.38 days / maximum 12 working hours a day for the same MCD value. Moreover, if we add to these results the amount of tree cuts for trees shorter than 5 meters we find ourselves in the state of validating H1. To rapidly get to such results we must directly change the URL (the source of Fig.1) as shown in the source of Fig.3 (after "begin=", "end=" and "threshold=" parts, the latter actually representing the MCD value taken into consideration).

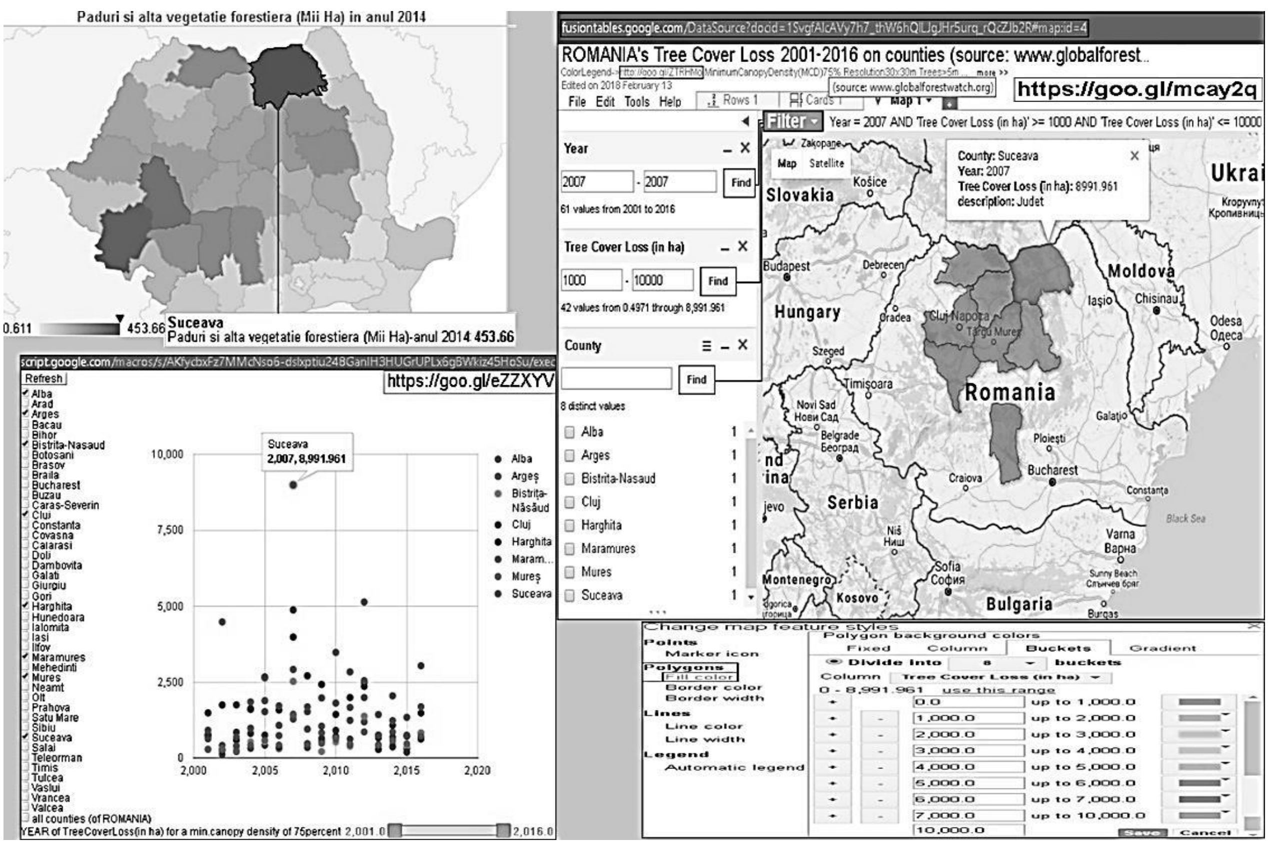
Based on such interactive representations, we compared again the GFW's query results with NIS's evidences for a period of time in which the GFW data was reported with both TCL and TCG (2001-2012). A total difference between TCL and TCG in favor of TCL was detected (net TCL) for that period, namely 64889 ha (218530 - 153641 for MCD=75%, Fig.1). When computing the difference between the NIS's TFA values for 2012 and 2001, respectively, a positive difference (total increase) of 162300 ha (6529100-6366800) is resulting in favor of 2012. Considering those different signs for the values computed above, we get a considerable aggregated absolute difference (227189 ha=|-64889| + |+162300|) for 2001-2012 between GFW and NIS's evolution reports. This difference means 3.57% from the TFA in 2000, almost three times (2.77 = 227189 / 81818) the initial calibration difference from the same year and only one and a half times the cumulated amount of artificial regenerated forest (1.47 = 227189 / 154334) as reported by NIS between 2001 and 2012. We assume the rest of 72855 ha (227189-154334) as young forest (trees under 5m) naturally grew after all type of cuts (including regeneration cutting with total values of 1018328 and 2053468 ha according to NIS for 2001-2012 and 1990-2016, respectively) and most likely not included in the GFW records for some years, until maturation. By this latter assumption, the slight overall increase of TFA is confirmed (H2).

Figure 3. Interactive online geo-representation - transparent parameters in URL and time progress bar / play button to indicate Romania's forest change (2001-2016)

Source: GFW, 2018c

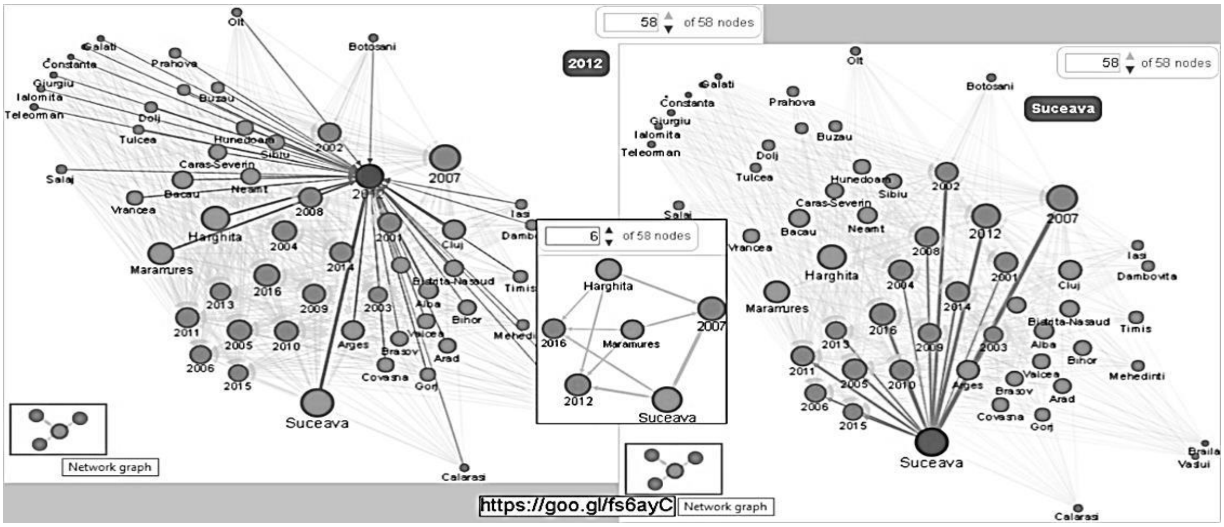
INS has online built-in representation facilities for forest vegetation such as heat maps in a geographic context. They come with basic interactions as immediate response with data for every single selected county and therefore support for understanding the whole structure. We can get more insight (e.g. the forestry richness on Romanian counties - the upper left side of Fig.4, is not always proportional to their forest loss, with few exceptions such as the Suceava county - the rest of Fig.4) when combining these representations with other interactive ones. For instance: (a) the ones involving forest loss data and behaving as online queryable maps such as Google Fusion Tables experimental data visualization project fed with satellite data (GFW, 2018d) first of all un-pivoted using Open Refine (Homocianu and Airinei, 2014b; Google Sheets, 2018d; Google Sheets, 2018e) and a custom KML-Keyhole Markup Language file for Romanian counties (filled with a certain color) and related polygons with geographic coordinates; (b) those acting as simple online charts with real-time control and feed-back such as custom representations in Google Apps Script using the Cartesian Point metaphor; (c) the ones made using the network graph option (**Figure 5**) in Google Fusion Tables and starting from the same un-pivoted data suggesting at the same time the counties and years with most severe TCL values: Suceava, Harghita, Maramureș, 2007, 2012 and 2016.

Figure 4. Queries on a NIS interactive map vs. a Google Fusion Tables queryable map and an interactive chart in Google Apps Script (GAS): counties with most forests vs. TCL



Source: NIS, 2014 – upper-left corner) and (Source: authors' work as online representations

Figure 5. Interactive online network graphs with TCL data (MCD=75%)

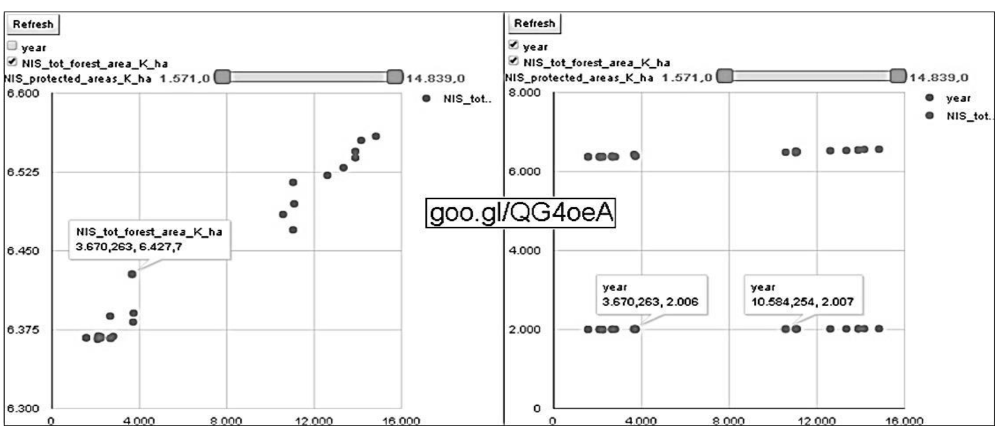


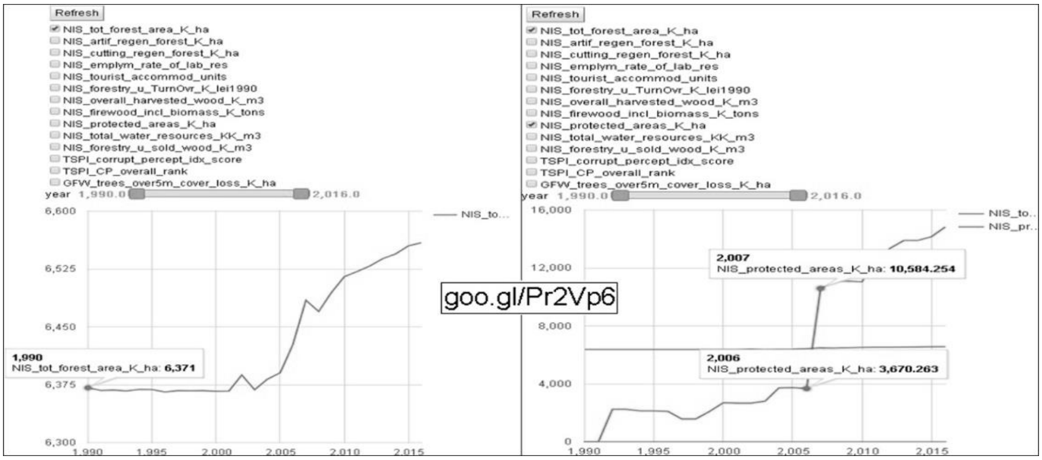
Source: authors' work and online representations using Google Fusion Tables

More, some other online geographic representations (Earth Engine Partners, 2013b) are able to capture both globally and for a specific area the forest extent and change (TCL, TCG) in last years (2000-2016) and provide suggestive insights by associating each year with a certain color (the GFC project).

The results obtained using the Excel DM add-in indicated ARIMA 0, 1, 1 for TFA with corresponding coefficients and equation proving a slight positive influence of the variable associated with TAU (**Figure 8**), already confirmed by the literature. Positive relations resulting from Excel DM have been also identified between TFA and PA or between TAU and PA (Fig.8), confirming most of the H2 hypothesis. Both negative and positive relations resulted between CRF and PA (Fig.8) depending on before or after mid-2006, most probably associated with Romania's EU pre-accession reforms that which included the declaration of additional protected areas (**Figure 6**).

Figure 6. Noticeable gaps in both the interdependence and evolution of Romania's total area with forests (1990-2016) and protected areas (1992-2016) as reported by NIS



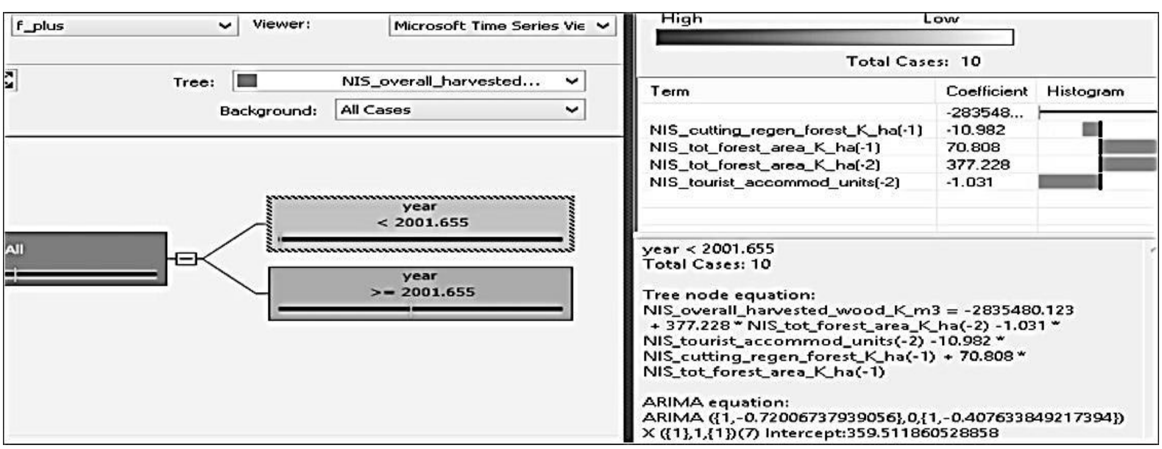


Source: authors' work and online interactive representations using Google Apps Script - GAS

The entire confirmation of H1 is based on admitting the positive relation identified between TCL and the accidental products cutting despite a relatively low coefficient (0.05 - as seen in Fig.8).

The resolution of the H3 hypothesis made us search for more data. In fact this new type of data characterizes the evolution of public property share on the forest fund including or not public administrative territorial units (an overall decrease and three sudden ones), in opposition to the one of the private property showing an increase (GAS, 2018f). For 2005-2016 the data (last 4 columns - only for representation reasons) was compiled and shared together with the full online data source (Google Sheets, 2018b) based on corresponding published government results (Ministry of the Environment, 2016). For 2004, a study on forest retrocession laws (Nichiforel, 2005) from July, 2005 served us as primary source. For 1990, 1996, 1998 and 2000 we have used as primary source for this type of data a scientific article from ten years ago. The remaining data (missing percentages) was regenerated by using formulas computing progressive averages together with a field indicating 0 for native and 1 for regenerated values. Even just for native data, the corresponding representations indicate sudden changes after 1990, 2000 and 2005.

Figure 7. Condition dependent ARIMA tree node equation as part of a persistent model suggesting exceptions for a positive relation OHW-CRF (Excel DM and SSAS)



Source: authors' work and a part of a data mining model

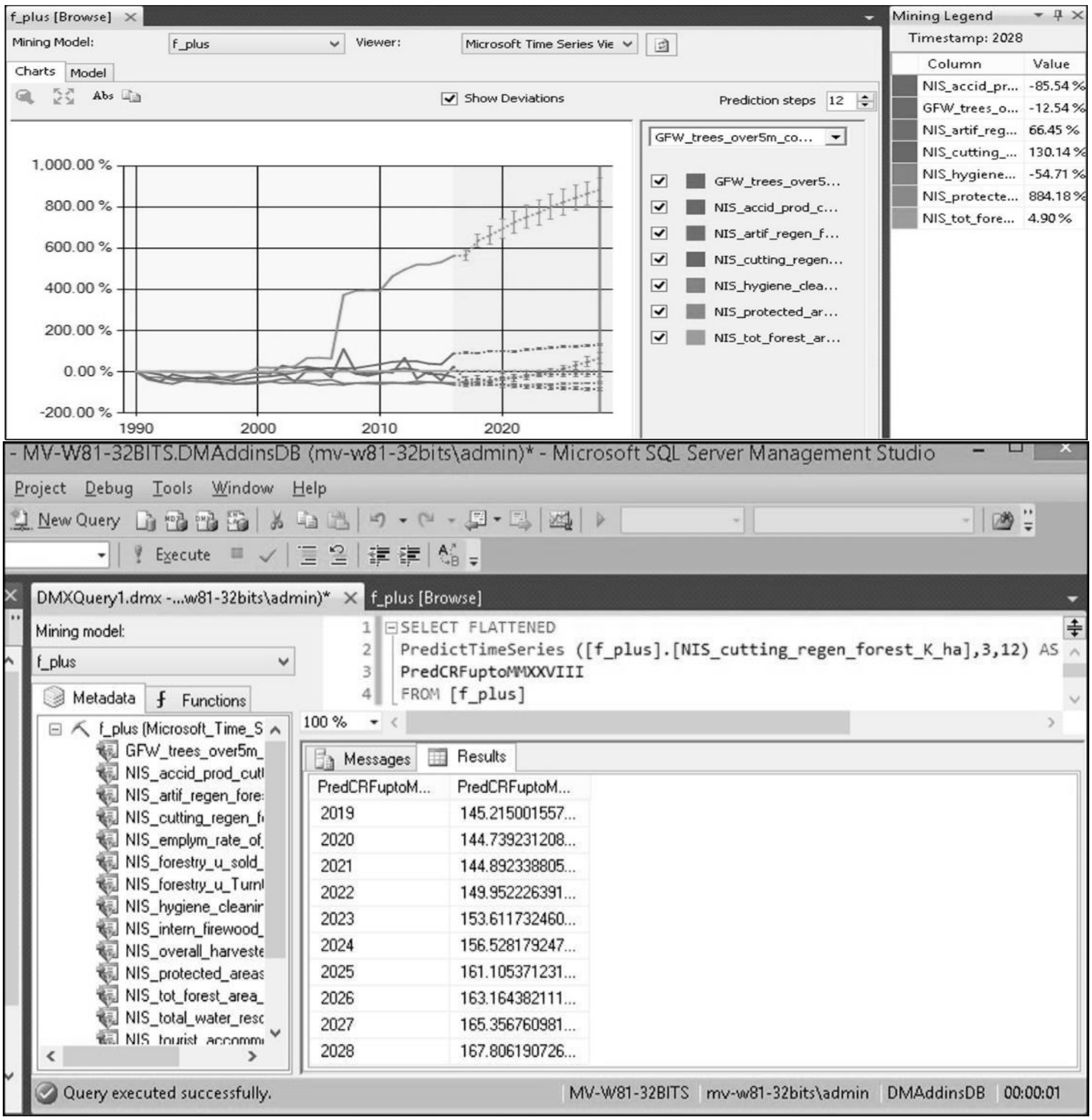
Figure 8. The metadata of the model obtained after using Excel DM

Variable to predict	MSOLAP node score	Tree split condition	Tree node equation	Type of ARIMA equation	Intercept	Total cases
NIS_tot_forest_area_K_ha or **TFA**	7.95	Not available	1948.513 +0.015 * TAU(-3) +0.020 * TAU(-1) -0.003 * PA(-1) +0.705 * TFA(-1) -0.026 * TFA(-2) + 0.0007 * TAU(-2)	(0,1,1)	7.231	24
NIS_artif_regen_forest_K_ha or **ARF**	1.98	Not available	27.624 -0.072 * CPOR(-5) -0.007 * APC(-8) -0.0004 * OHW(-5)	(1,2,1)	0.418	15
NIS_cutting_regen_forest_K_ha or **CRF**	4.7	Not available	1428.309 +0.004 * TAU(-1) -0.208 * TFA(-1) +0.022 * TAU(-4) +0.003 * TAU(-3) -1.724 * ERLR(-7)	(1,1,1)	4.382	20
NIS_hygiene_cleaning_cutting_K_ha or **HCC**	11.28	year <2003.5	342.604 + 0.798 * ARF(-1) -0.00001 * TO(-1) +0.562 * HCC(-1)	(2,0,8)	189.127	13
		year >=2003.5	789.074 +0.305 * HCC(-1) +4.062 * ERLR(-1) -0.00009 * TO(-1) -14.046 * ARF(-1)	(2,0,8)		13
NIS_accid_prod_cutting_K_ha or **APC**	9.55	ERLR-1 < 63.05	-85.608 -0.189 * HCC(-4) +12.521 * ERLR(-1)	(1,0,1)	0	12
		ERLR-1 >= 63.05	1516.708 -18.724 * ERLR(-1) +0.198 * HCC(-4)	(1,0,1)		11
NIS_emplym_rate_of_lab_res or **ERLR**	4.54	Not available	15.449 +0.008 * HCC(-1) +0.669 * ERLR(-1)	(0,1,1)	-0.608	26
NIS_tourist_accommod_units or **TAU**	14	Not available	27177.076 -0.114 * TAU(-4) +12.065 * CRF(-2) +0.696 * TAU(-1) +0.220 * TAU(-2) -4.417 * TFA(-1) +0.347 * TAU(-3)	(0,1,1)	143.577	23
NIS_forestry_u_TurnOvr_K_lei 1990 or **TO**	26.66	Not available	6871443.799 -66466.308 * ERLR(-3) +0.412 * TO(-1)	(1,0,3)	608558.5	24
NIS_overall_harvested_wood_K_m3 or **OHW**	14.48	year <2001.655	-2835480.123 +377.228 * TFA(-2) -1.031 * TAU(-2) -10.982 * CRF(-1) +70.808 * TFA(-1)	(1,0,1)	359.512	10
		year >=2001.655	-199318.133 +30.471 * CRF(-1) +22.052 * TFA(-1) -2.121 * TAU(-2) +12.566 * TFA(-2)	(0,1,0)(7)		15
NIS_intern_firewood_biomass_K_to or **IFB**	18.34	Not available	5788.509 +0.0009 * TO(-8) +48.541 * CPOR(-5) -40.316 * ERLR(-8) +53.712 * CPOR(-4)	(0,0,1)	12686.32	15
NIS_protected_areas_K_ha or **PA**	16.17	year <2006.753	-191675.925 +32.586 * TFA(-2) -24.568 * CRF(-2) +0.298 * TAU(-1) +0.681 * TAU(-2) +0.740 * PA(-1) -2.654 * TFA(-1)	(0,1,1)	524.94	14
		year >=2006.753	-244511.444 -0.166 * TAU(-2) -0.251 * PA(-1) +26.590 * TFA(-1) +12.539 * TFA(-2) +0.399 * TAU(-1) +44.840 * CRF(-2)	(0,1,1)		10
NIS_total_water_resources_KK_m3 or **TWR**	12.93	Not available	69909.777 -0.049 * PA(-6) -4.370 * TFA(-6) -637.311 * CPI(-2)	(1,0,1)	10025.05	18
NIS_forestry_u_sold_wood_K_m3 or **SW**	15.34	Not available	20593.729 -0.145 * PA(-7) -0.255 * OHW(-5) -45.088 * CPOR(-7)	(1,0,1)	0	13
TSPI_corrupt_percept_idx_score or **CPI**	-0.25	Not available	1.139 +0.008 * CRF(-4) +0.010 * CRF(-5) +0.0003 * TAU(-7) +0.003 * CRF(-3)	(1,0,6)	4.713	20
TSPI_CP_overall_rank or **CPOR**	3.89	Not available	245.792 -2.118 * ERLR(-1) -0.002 * IFB(-2) -3.012 * CPI(-3)	(1,0,1)	26.77	17
GFW_trees_over5m_cover_loss_K_ha or **TCL**	3.79	Not available	21.399 +0.002 * IFB(-2) +0.050 * APC(-5) -0.004 * OHW(-6)	(1,0,1)	8.659	14

Source: authors' own work and the resulting synthesis of a data mining model

In addition, the results from Excel DM indicate two ARIMA specifications (Figs.7 and 8) for the overall amount of harvested wood depending on a particular moment, namely before or after mid-2001 (confirmation of H3). These two demonstrate the existence of certain specificities in the context of an increasing private property share by a changing sign of the dependence on the areas covered by regeneration cutting which tends to increase considerably as shown in Fig. 9 which captures the results of a DMX-Data Mining eXtension query (OHW-CRF - usually a positive relation).

Figure 9. Forecast settings and a DMX query generating both growth rates and absolute values over the next 12 years and for the last 10 (as 12-3+1) of the next 12



Source: authors' work

To validate the H4 hypothesis it is worth mentioning the coefficients with different signs corresponding to the relations between the overall amount of internally produced firewood including biomass as primary energy resource and the corruption perception overall rank (IFB-CPOR, Fig.8) or the overall employment rate of labor resources (IFB-ERLR, Fig.8). Moreover the amount of turnover for forestry units with an overall ascendant trend seems to be negatively influenced by the aforementioned rate (TO-ERLR, Fig.8) as a generic indicator of economic health.

Conclusions

The papers' main contribution is that of creating a support model for validating many hypotheses of a real interest for both the general public and the academic community, concerning the evolution of the forest fund and the corresponding variables that have proved to be important.

The importance of this model essentially based on data mining techniques and acting as a queryable knowledge piece is coupled with the practical nature of all efforts including at least three key elements: the forest loss data for almost two decades and forty counties transformed into an interactive online map with enriched query facilities including minimum and maximum forest loss thresholds; compiled data about various indicators made available as web resources together with their interactive online representations as evolution in time and correlation charts; direct and indirect references to online data quantification, correction, transformation and integration examples also made available for general use.

With no doubts regarding still existing theoretical and practical limitations for this study mainly due to variables not yet included in the analyses and few or missing data, we hope of being able to provide further improvements and considerations on this challenging theme.

Acknowledgments

We would like to recognize the online support of our data providers: The Romanian Court of Accounts, The Romanian Ministry of the Environment, The National Bank of Romania, The National Institute of Statistics from Romania, The Global Forest Watch and The Transparency International - Global Coalition against Corruption. We would also like to confirm Google's generous contribution with online tools such as: Google Drive, Google Fusion Tables experimental project and Google Apps Script as self-contained environment for code editing in a browser. Finally we would like to thank for the consistent support of our provider of application development environments and data storage and analysis solutions, Microsoft, through the Imagine (formerly DreamSpark) academic software licensing program.

List of acronyms and abbreviations

ARIMA – Auto-Regressive Integrated Moving Average

ARTXP – Auto Regressive Trees with Cross Predict

BNR – National Bank of Romania

DM – Data Mining

DMX – Data Mining eXtensions to SQL

DNA – National Anti-corruption Department / Division

EU – European Union

EVI – Enhanced Vegetation Index

FAO – Food and Agriculture Organization of the United Nations

GAS – Google Apps Script

GFC – Global Forest Change

GFW – Global Forest Watch

GFW_trees_over5m_cover_loss_K_ha / TCL – Tree Cover Loss

H1, H2, H3 and H4 – Hypotheses No.1, 2, 3 and 4

IPC / ACPI – Annual Consumer Price Index

K ha – Thousands hectares

K m3 – Thousands of cubic meters

KK m3 – Millions of cubic meters

KML – eXtensible Keyhole Markup Language for expressing geographic annotations

MCD – Minimum Canopy Density

NASA – National Aeronautics and Space Administration

NDVI – Normalized Difference Vegetation Index

NIS / INSSE – National Institute of Statistics from Romania

NIS_accid_prod_cutting_K_ha / APC – Forest area where cuts of accidental products occurred

NIS_artif_regen_forest_K_ha / ARF – Forest area where artificial regeneration took place

NIS_cutting_regen_forest_K_ha / CRF – Forest area where regeneration cutting occurred

NIS_emplym_rate_of_lab_res / ERLR – Employment rate of the labor resources

NIS_forestry_u_TurnOvr_K_lei1990 / TO – Turnover for forestry units

NIS_hygiene_cleaning_cutting_K_ha / HCC – Forest area where hygiene and cleaning cutting occurred

NIS_intern_firewood_biomass_K_to / IFB – Overall amount of firewood including biomass as primary energy resource

NIS_overall_harvested_wood_K_m3 / OHW – Overall amount of harvested wood

NIS_protected_areas_K_ha / PA – Protected areas

NIS_tot_forest_area_K_ha / TFA – Total Forest Area

NIS_total_water_resources_KK_m3 / TWR – Total water resources

NIS_tourist_accommod_units / TAU – Number of tourist accommodation units

No. – Number

OLAP – On-Line Analytical Processing

PNA – National Anti-corruption Prosecutor's Office

ROL – ROmanian Leu

RON – ROmanian New Leu

SQL – Structured Query Language

SSAS – Microsoft SQL Server Analysis Services

TC – Tree Cover

TCG – Tree Cover Gain

TJ – Terra Joules

TOA – Top of Atmosphere

TSPI – Transparency International - Global Coalition against Corruption

TSPI_corrupt_percept_idx_score / CPI – Corruption perception index (score)

TSPI_CP_overall_rank / CPOR – Corruption perception overall rank

URL – Uniform Resource Locator

USGS – United States Geological Survey

REFERENCES

1. Bailey, P., (2007), Think of an Elephant: Combining Science and Spirituality for a Better Life, London: Watkins Publishing.
2. Boone, L.E., Kurtz, D.L., (2014), Essentials of Contemporary Business, Binder Ready Version, 1st Edition, Hoboken, John Wiley & Sons.
3. Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., (2015), Time Series Analysis: Forecasting and Control. 5th ed., Hoboken, NJ, Wiley.
4. Card, S.K., Mackinlay, J.D., Shneiderman, B., (1999), Readings in Information Visualization: Using Vision to Think, San Francisco, Morgan Kaufmann Publishers.
5. Chapman, B., (2011), Geopolitics: A Guide to the Issues, Santa Barbara, Praeger.
6. Chouinard, Y., Ellison, J., Ridgeway, R., (2011), The Sustainable Economy, Harvard Business Review, the October issue.
7. Dal Bó, E., Hernández, P., Mazzuca, S., (2015), The paradox of civilization: pre-institutional sources of security and prosperity, NBER working paper series.
8. Fodor, A., Pică, E.M., (2012), Comparative study of legislation in E.U. and Romania on protected areas, Scientific Annals of the Constantin Bracusi University of Targu-Jiu, Engineering Series, 3, 233-245.
9. Gabryjończyk, P., Kułaga, K., (2017), Relationship between tourist accommodation establishments and natural environment in Eastern Region of Poland, World Scientific News, 78(1), 376-382.
10. Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., (2013), High-resolution global maps of 21st-century forest cover change, Science, 342(6160), 850-853.
11. Homocianu, D., Airinei, D., (2014a), Business Intelligence Facilities with Applications in Audit and Financial Reporting, Audit Financiar, 117(9), 17-29.

12. Homocianu, D., Airinei, D., (2014b), Consolidating Source Data in Audit Reports, Audit Financiar, 116(8), 10-19.
13. Meek, C., Chickering, D.M., Heckerman, D., (2002), Autoregressive Tree Models for Time-Series Analysis, Proceedings of the 2002 SIAM International Conference on Data Mining.
14. Mossay, E., De Kemmteter, M., (2017), Shifting Economy. Your Practical Toolkit to enter the Next Economies, Walhain-Perbais, Paperland.
15. Naish, J., Burns, A., (2017), Minimising operator dose during JET shutdown using virtual, Fusion Engineering and Design, 124 (11), 1215-1218.
16. Nichiforel, G., (2005), The state of implementation of the restitution laws on forests, Bucovina Forestieră, 13(1-2), 23-43.
17. Skaf, R., (2011), Survival: The Ultimate Mission, Boston, eBookIt.com.
18. Solum, L.B., (2013), Virtue Jurisprudence: Towards an Aretaic Theory of Law, In: Huppes-Cluysenaer, L., Coelho, N.M.M.S. (Eds.). Aristotle and the Philosophy of Law: Theory, Practice and Justice, Dordrech, Springer, 1-31.
19. Timmons, D., Harris, J.M., Roach, B., (2014), The Economics of Renewable Energy, Medford, Global Development and Environment Institute, Tufts University.

Web references

BNR, (2014), *The Leu's history*, Online at: http://www.bnr.ro/files/muzeu/2016/resurse/BNR_Istoria_Leului_ro.pdf

Business Dictionary, (2018), *Economic Sustainability*, Online at: http://www.businessdictionary.com/definition/economic-sustainability.html

Court of Accounts, (2013), *Synthesis of the Audit Report of the Romanian Court of Accounts on "The patrimonial condition of the forestry of Romania between 1990 and 2012"*, Online at: http://www.curteadeconturi.ro/Publicatii/economie7.pdf

EU Publications, (2015), *Sustainable agriculture, forestry and fisheries in the bioeconomy. A challenge for Europe: 4th SCAR foresight exercise - Study,* Online at: https://publications.europa.eu/en/publication-detail/-/publication/7869030d-6d05-11e5-9317-01aa75ed71a1

Earth Engine Partners, (2013a), *Global Forest Change 2000-2016. Data Download*, Online at: http://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.4.html

Earth Engine Partners, (2013b), *Global Forest Change*, Online at: http://earthenginepartners.appspot.com/science-2013-global-forest

Elsevier, (2018), *iPlots Authoring Tool*, Online at: http://authortools.elsevier.com/interactive plots/verification

FAO, (2015), *FRA 2015. Terms and definitions*, *Forest Resources Assessment Working Paper 180*, Online at: http://www.fao.org/docrep/017/ap862e/ap862e00.pdf

GAS, (2018a), *Interactive evolution* (*enable only the 1st and 11th variable -both/one at a time)*, Online at: https://goo.gl/Pr2Vp6

GAS, (2018b), *Interactive corellogram* (*enable those 2 variables - both / one at a time)*, Online at: https://goo.gl/QG4oeA

GAS, (2018c), *Interactive corellogram (enable those 2 variables - both / one at a time)*, Online at: https://goo.gl/5HgS6F

GAS, (2018d), *Interactive evolution (enable only the 3rd and 9th variable -both/one at a time)*, Online at: https://goo.gl/Pr2Vp6

GAS, (2018e), *Interactive evolution (enable only the last variable)*, Online at: https://goo.gl/Pr2Vp6

GAS, (2018f), *Interactive evolution (enable only last four variables at once)*, Online at: https://goo.gl/Pr2Vp6

GFW, (2018a), *Forest change*, Online at: http://www.globalforestwatch.org/map/17/47.18/27.57/ROU/grayscale/loss,forestgain,forest2000?tab=analysis-tab&begin=2001-01-01&end=2013-01-01

GFW, (2018b), *Forest change*, Online at: http://www.globalforestwatch.org/map/17/47.18/27.57/ROU/grayscale/loss,forestgain,forest2000?tab=analysis-tab&begin=2001-01-01&end=2013-01-01&threshold=30

GFW, (2018c), *Forest change*, Online at: http://www.globalforestwatch.org/map/7/45.48/25.36/ROU/grayscale/loss,forestgain?tab=analysis-tab&begin=2001-01-01&end=2017-01-01&threshold=75

GFW, (2018d), *Download the country data for Romania*, Online at: http://www.globalforestwatch.org/country/ROU

Google Earth Engine, (2018), *Datasets*, Online at: https://earthengine.google.com/datasets/

Google Sheets, (2018a), *Forest fund's equation for a linear trendline starting from 1800*, Online at: https://goo.gl/jPSVqc

Google Sheets, (2018b), *Extended dataset as online spreadsheet - 0s instead of NULLs for representation reasons*, Online at: https://goo.gl/fKJ3v9

Google Sheets, (2018c), *Average of working days between 2001 and 2016 using the NETWORKDAYS formula*, Online at: https://goo.gl/m3kp67

Google Sheets, (2018d), *RO-Tree-Cover-Loss2001-2016counties-min75%-canopy-density (multidimensional)*, Online at: https://goo.gl/n1kgUv

Google Sheets, (2018e), *RO-Tree-Cover-Loss2001-2016counties-min75%-canopy-density (relational)*, Online at: https://goo.gl/ioLMMv

Green Peace, (2015), *Forest loss map*, Online at: http://www.greenpeace.org/romania/ro/campanii/pa-duri/publicatii/harta-pierderii-padurii-3-hectare/

Green Peace, (2017), *Report: Illegal cutting from the forestry of Romania in 2016*, Online at: http://www.greenpeace.org/romania/Global/romania/paduri/raport-taieri-ilegale-2016.pdf

Holzindustrie Schweighofer, (2014), *Holzindustrie Schweighofer - Austrian investor of the year in Romania*, Online at: https://www.schweighofer.at/fileadmin/files/all_en/Press/Holzindustrie_Schweighofer_Investor_of_the_year_in_Romania.pdf

Johnson, J.E., Smith, D.W., Short, S., (2009), *Principles of Regeneration Silviculture in Virginia*, Online at: https://pubs.ext.vt.edu/content/dam/pubs_ext_vt_edu/420/420-405/420-405_pdf.pdf

Microsoft, (2017), *Microsoft Time Series Algorithm Technical Reference*, Online at: https://docs.microsoft.com/en-us/sql/analysis-services/data-mining/microsoft-time-series-algorithm-technical-reference

Ministry of the Environment, (2016), *State of forest*, Online at: http://www.mmediu.ro/categorie/starea-padurilor/209

Ministry of Waters and Forests, (2017), *We will provide about 4 million cubic meters of firewood to the population*, Online at: http://apepaduri.gov.ro/doina-pana-lemnul-de-foc-din-padurile-statului-trebuie-sa-ajunga-la-populatie-nu-sa-ingrase-niste-conturi/

Nichiforel, L., (2014), *Forestry course*, Online at: http://silvic.usv.ro/cursuri/silvotehnica_id.pdf

NIS, (2011), *Energy consumption in households in 2009*, Online at: http://www.insse.ro/cms/files/publicatii/CENG_publicatie_tabele.pdf

NIS, (2007), *Romanian Statistical Yearbook. Agriculture and sylviculture*, Online at: http://www.insse.ro/cms/files/pdf/en/cp14.pdf

NIS, (2014), *Statistical Atlas. Forests and other forest vegetation (Thousand Ha) in 2014*, Online at: http://www.insse.ro/cms/ro/content/atlas-statistic

NIS, (2017), *ACPI - data series*, Online at: http://www.insse.ro/cms/ro/content/ipc-serii-de-date