Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

Beyond big data – new techniques for forecasting elections using stochastic models with self-organisation and memory

Dmitry Zhukov [a], Tatiana Khvatova [b,c,*], Carla Millar [d], Elena Andrianova [a]

[a] *Russian Technological University (RTU MIREA), 78, Vernadskogo Str., Moscow, 119454, Russian Federation*
[b] *EMLyon business school, Entrepreneurship & Innovation Research Center, France, 23 Avenue Guy de Collongue, Écully, 69130 France*
[c] *Lappeenranta - Lahti University of Technology LUT, Yliopistonkatu 34, Lappeenranta, 53850 Finland*
[d] *University of Twente, Drienerlolaan 5, NB Enschede, 7522 Netherlands*

A B S T R A C T

This paper introduces an innovative social process model addressing population-wide measures of voter preferences that was tested on data from the 2016 US presidential election. Population-wide, "macroscopic" parameters are needed when privacy, ethics or regulatory constraints block "big data" techniques (e.g., in political contexts to counter "micro-targeting"). Confidence will be eroded if existing trend models and other macroscopic approaches frequently fail to predict outcomes, however campaign data reveal mathematical features that suggest a different possible approach. Given that the populations modelled exhibit self-organisation and memory when transmitting viewpoints, our model is based on mathematical representations of such processes. Its validation indicates the applicability and potential generalisability of this theoretical approach. In order to design a stochastic dynamics model of changing voter preferences, we evaluated probability models for transitions between possible system states (magnitudes of voter preferences), formulated the boundary task for probability density functions and derived a second-order non-linear differential equation incorporating self-organisation and memory. We find consistent dependencies between influences on the system and its reaction, and it is congruent with empirical data. The ability to use researchable global parameters indicates the potential for modelling electoral processes and wider applicability for complex social processes, avoiding dependence on "internal" variables.

1. Introduction

Although "big data" opens rich new seams of investigation and analysis, there are areas where its deployment may be constrained, despite confidence in reliable forecasting being crucial for both political and economic action. Forecasting electoral outcomes is a classic case, and the challenge is to improve forecasting using macroscopic techniques, i.e., working with global totals rather than being dependant on the analysis of big data. When these global techniques 1.are complemented by more fine-grained approaches, much can be learned, but for many practical purposes – including campaign management, political risk assessment and investment strategy – having forecasts based on the progression of global totals is a priority. Such forecasts can also provide a template for evaluating big data results: if the big data analysis fails to match the observed global trend, then questions naturally arise about the quality of the datasets and/or the algorithms. If a reliable global forecasting method exists and its results match those of the "big data" approach, this can provide users with initial confidence in the power of the latter.

The 2016 US presidential election may be an example of this: uncertainty arose regarding the impact of fake accounts that were created on social networks to organise information attacks on users with the aim of influencing their electoral preferences. Such activities and the difficulty of detecting and assessing them at a micro-level may undermine confidence in democratic electoral institutions and may ultimately have a negative influence on the economy. In the long run, investor behaviour is influenced by possible political risks connected with the victory of a particular candidate or party as this may lead to changes in the tax regime, customs duties, various support programmes, etc. State authorities, investors and business people require reliable forecasts of electoral processes to be able to make long-term investments and returns. It is vital that these parties know that there are no political risks

* Correspodning author.
E-mail address: khvatova@em-lyon.com (T. Khvatova).

connected with illegal influence on electoral processes, which can lead to unpredictable consequences. Developing innovative technological models that address the need to forecast election results, even in the presence of part-trackable or unrecognised micro-targeting and influence, is therefore very important in order to provide confidence in the system and ensure the stability of economic processes.

Electoral processes have very particular characteristics because they occur within complex social systems, and this should be taken into account when analysing and modelling them. At any moment of time, there are numerous explicit and hidden social and economic factors influencing the dynamics of voters' preferences. These factors are often conflicting, and may strengthen or weaken one another. The net effect is that at the micro level, everyone makes their own decisions based on their individual goals and expectations. However, history (personal, group and society-wide) and interaction with others are recognised by political analysts as crucial to understanding the development of swings in electoral preference. Powerful as "big data" tracking may be, there are limits to the acceptability of tracking an individual's political stance and the content of discussions they may have with others during a campaign.

That is why, in the authors' view, considering electoral dynamics at a statistical level, i.e., not considering one person, but millions and tens of millions of people, means that we can describe these dynamics using the language of "big data in probability theory". Adding multiple divergent "movements" of people forms a flow that can define the direction of the electoral process. The action of one individual can be considered random, but the sum of these movements as a whole becomes a trend. We emphasise again that if we do not look at the process on a "big numbers" level, then it is clear that the "movement" of each individual is caused by certain circumstances. The divergent directions of these "movements" nonetheless lead to statistical averaging on the macro level.

On the macro level, the state of the complex social system in which the electoral process takes place is expressed as the proportion of voters ready to endorse a particular candidate or party at a certain moment in time. On the micro level, voters can change their preferences when influenced by obvious or hidden social and economic factors. Revealing those factors is a sophisticated, and time-consuming, sociological task that involves preparing surveys, building a representative sample, analysing results, etc. The various respondents may be influenced by different factors that are difficult to capture when preparing a sociological survey, but – for psephological purposes – the macro phenomenon of changes in the proportion of voters in the respective "camps" is the dependant variable and predicting it is the criterion of success in discovering relevant factors. This necessitates a sociological approach, although it should be noted that any attempt to provide valuable insights and explanations may find that additional – unresearched – factors are needed to offer a complete explanation.

At the same time, on the macro level, changes in electoral preferences are reflected through measures such as opinion polls that can be followed in real time. The resulting time series can be analysed and used to model behaviour. This offers an obvious practical technological advantage compared to sociological survey approaches. The task in hand can be compared to modelling the behaviour of stock and raw material markets in which multiple divergent factors operate, and where their totality defines the direction of changes.

In the authors' view, developing new models designed to forecast the dynamics of electoral processes based on statistical macro approaches and probability theory methods could be useful for the wider circle of economics and sociology scientists.

In our paper, we consider statistical data relating to the 2016 US presidential election campaign. It is important to note that this dataset was selected because it is publicly available and of high quality. Using this data as a validation set, we attempted to develop a generalised model that accounts for self-organisation within the system and memory with respect to its previous states. The model is potentially universal and applicable to any electoral process.

Social systems are very different from technical ones due to the active influence of the human factor. The human factor introduces uncertainty regarding the effects of ongoing influences on systems, which in turn creates stochastic behaviour and can also be reflected in self-organising effects in systems. We also need to take into account the potential for memory regarding previous states of the system, and the previous influences it has been exposed to. As a result, the dynamics of processes are significantly non-linear in nature.

To model the dynamics of self-organising social systems with memory, we need to apply the methods and resources of theoretical computer science and cybernetics, in particular the science of controlling both technical and biological (social) systems as defined by Norbert Wiener (1914, 1954). Using the common principles and methods of theoretical computer science to model technical and informational systems will arguably enable us to obtain high-quality results for modelling social, economic and socio-technical systems.

This paper describes the results of our research into the stochastic dynamics of self-organising systems with memory, based on the example of describing electoral campaigns. The goals of the research are: a) to conduct statistical analyses of electoral processes and reveal the role of self-organisation and memory in the mechanisms of electoral shifts, and b) by analysing electoral campaigns, to create a quantitative model that describes the essentially non-linear dependency between the campaign's impact on voters and shifts in their preferences.

We adopted the following approach in our study. We examined accurate data for a social process (the development of voter preferences during an electoral campaign), which we then processed to obtain histograms and statistical characteristics of the parameters of observed shifts and their dependencies. Based on this analysis, and accounting for these specific characteristics, we then developed a dynamic model for the social process observed. By studying the dynamics of the electoral process as a social process, we address both statistical analysis and the modelling of stochastic processes.

Forecasting electoral campaign results is very important in terms of overcoming social uncertainty, which has a significant influence on the economy. The proposed model develops an algorithm for electoral campaign forecasting and decision-making. The algorithm is described in Section 6, "An algorithm for forecasting electoral campaign results and decision-making". Election forecasts can influence numerous important economic processes. First, forecasting a candidate's victory allows investors to make long-term decisions about investing in various sectors of the economy based on the political and economic views of the candidates or parties participating in the elections. Secondly, if the victory (or defeat) of a candidate or party no longer carries a risk of surprise, this will have a positive effect on stock index dynamics (there will be no sharp jumps in prices, no sudden defaults, etc.). Thirdly, forecasting political trends will have a positive effect on countries' opportunities to issue bonds and attract financing at lower interest rates.

To clarify our research scope, we note that this paper does not attempt to explore social behaviour or the reasons behind voting intention. It instead provides a framework for systematically representing (modelling) the development of voting intention.

We demonstrate the superiority of our approach, taking into account memory and self-organisation, when choosing a mathematical representation of how an electoral process develops. A trivial representation, for example, would be a "straight line" from the initial level of support to the final level. More sophisticated representations arise when non-linearity is observed in polls during a campaign. The mathematical techniques for representing the transition from one level to the next assume that this transition is not a series of straight-line jumps from one value to the next, i.e., it is not a jagged line. Instead, they assume that there is a process of development that is best shown by a curvilinear representation. Theoretically, there are many mathematical representations that might match the fluctuation patterns observed. The choice of model has, in the past, been guided by an interpretation that considers the fluctuations to be the result of influences working discretely on each

individual, and possibly by simple transmission from individual to individual. Such approaches succeed in producing mathematical representations in which the transition from one observed level to the next is a smooth curve. However, they are far from successful in representing the whole series of observations throughout a campaign. The starting point of our work is the insight that problems with existing mathematical formulations arise – at least in part – because there is more at work in an electorate than simple discrete influences and one-off transmissions from one individual to another. We therefore focus on a mathematical approach that is appropriate to self-organising networks with memory, and investigate whether it provides a better tool for representing the non-linear nature of preference development in an electorate. Since the representation is intended to have general application and to retain its congruence with more general network modelling, we do not seek to explore or explain the nature of the influences that are at work. We do, however, demonstrate that this modelling approach can be used to achieve a mapping onto real-life data that is superior to the results generated using other methods, and that this mapping allows the influences on voter sentiment to be represented by parameters. The factors that play a role in social situations are by nature complex and varied, so we accept that representing them in the form of single positive and negative parameters is a simplification. All modelling, however, is a process of simplification, and models can only be useful and understandable if they show what is modelled in terms of some independently derived formula, pattern or process.

We hope that this paper provides people studying electoral processes with a mathematical formulation that allows them to look at data differently and highlight where to seek causes and how to track the evolution of voting intentions.

We start with an overview of existing research into the analysis and modelling of social processes and electoral processes in particular (Section 2). In Section 3, we present the data selected for our modelling (data on the 2016 US presidential election campaign) and describe the algorithm that will be used to analyse the data. In Section 4, we analyse the data, laying the foundations for our new model on the dynamics of electoral processes. Our proposed model of the stochastic dynamics of states in a self-organising social system with the potential for memory is discussed in detail in Section 5. First, we develop probabilistic difference schemas of transitions between states (Section 5.1); secondly, after formulating and solving the boundary value problem, we obtain the primary equation of the model (Section 5.2); finally, in Section 5.3, we present the modelling results and explain our new model of voter preference shifts, taking into account self-organisation and memory. In Section 6, we propose an algorithm for forecasting electoral campaign results that can be used to assist decision-making. The paper closes with our conclusions.

2. An overview of existing research on the analysis and modelling of social processes

Social processes are characterised by complex mechanisms and stochastic behaviour, where multiple, different states depend on the effects of participants interacting with one another (Easley et al., 2010; Karsai et al., 2014; Centola and Macy, 2007; Andrianova et al., 2020). The theoretical approach used to describe such social systems has a lot in common with the kinetic description of physical systems (Barrat et al., 2008; Watts, 2002; Gleeson and Cahalane, 2007), which helps greatly in understanding their behaviour. For example, kinetic models using threshold mechanisms primarily consider cascade or chain phenomena. These models consider a social system or a network (nodes) in which a macroscopic share of users who have the same specific status emerges rapidly from a specific microscopic state through the capture of many new nodes. This approach was created in early theories (Granovetter and Soong, 1983) and is described in detail in the global cascades model (Watts, 2002). The significance of this model is indisputable (Watts, 2002; Gleeson, 2013, 2011; Yağan and Gligor, 2012; Nematzadeh et al.,

2014; Singh et al., 2013; Piedrahita et al., 2013), however observations of real processes in social networks show its limitations. In particular, the model focuses on instantaneous global cascades initiated by single local disturbances, despite the fact that threshold mechanisms in social systems may actually be more significant in some cases. Furthermore, the state of nodes may depend on external impulses coming from mass media or advertisements (Kocsis and Kun, 2011) and on the stochastic component of processes.

In some cases, in order to describe processes in social networks, stochastic approaches are used to consider the dependencies of states on time (Airoldi et al., 2008; Zhukov et al., 2017; Khvatova et al., 2017, 2020). For example, Airoldi et al. (2008) discuss the mixed-membership model in stochastically formed groups. Their model aims to explain pairwise changes such as the presence or absence of links amongst pairs of objects. The analysis of probabilistic changes amongst the pairs of objects requires special assumptions such as independence and the variability of the link (of mixed membership in stochastically formed groups). The model makes specific assumptions in order to track the dynamics of changes in the number of members in the emerging groups and the clustering of members by groups.

Previous research (Khvatova et al., 2017, 2016; Zhukov et al., 2018, 2021) has explored group social processes using percolation theory, which takes the network structure into account. The authors consider the influence of the average number of links per node in a social network on its percolation threshold (network conductivity in general) and the dynamics of its development.

The theory of multi-antenna systems can be used to model voter behaviour. Based on transition rules, agents take specific states, form a group of members who are linked by the properties of the group, and cooperate to solve tasks or reach a specific goal (Plikynas et al., 2015). In this model, the behaviour of agents, for example, or changes in voter preferences, can depend on dynamically changing conditions (Hay and Flynn, 2016).

The theory of cellular automata (Sigov et al., 2017; Wang et al., 2014; Li et al., 2013) can be applied to describe electoral processes and voter preferences. In Wang et al. (2014), the behaviour of the social system depends on the properties of the external environment and behaviour structure as described by four parameters: diversity, coherence, mutual dependency and adaptability. When mutual dependency and adaptability are increased, agents' behaviour becomes more organised and targeted.

Li et al. (2013) use cellular automata to model the evolutionary dynamics of social networks, studying the stochastic characteristics of clustering processes (partition into poorly connected groups), such as the degree of partition. Three factors necessary for building a social network are explored: selfishness, mutuality and altruism. The effects of these factors on network growth are also analysed.

It should be noted that most campaign research on social processes in elections is dedicated to important but very specific aspects (party manifestos, candidate performance, etc.). For example, (Bonica et al., 2014) describe the change in voter preferences (polarisation of views) amongst medical doctors in the US during the Federal Election campaigns from 1991 through to 2012, and how these physicians' financial contributions to the Republicans and Democrats correlated with both their gender and medical specialisation. The general conclusion provided by the authors (Bonica et al., 2014) was that between 1991 and 2012, the political preferences of American doctors shifted from Republican to Democrat. However, the reasons for this shift were not investigated.

Another important strand of research relates to how the information space (TV, newspapers, internet, social networks, etc.) fundamentally affects electoral preferences. Mundim (2012), for example, studied the influence of the mass media on voter preferences during the 2002 and 2006 Brazilian presidential elections. That research explored the following factors: the general party propaganda of each candidate, the time candidates spent making open public speeches and coverage of

candidates' activities by news outlets. Mundim concluded that party propaganda and candidate speeches had a greater impact on election results in Brazil than media coverage.

An overview of voter preference models is presented in the works of Dewan and Shepsle (2011), in which two classes of models are discussed: a) models in which elections are analysed by aggregating preferences, and b) models in which elections are analysed using information aggregation mechanisms. The models in the first category (a) usually consider general changes in the share of voters with certain preferences over time (i.e., they are phenomenological and do not associate the shifts with the reasons behind them). The models in the second category (b) consider the influence of the information provided about the candidates and parties on the voters during the elections (i.e., they partly associate the shifts with the reasons behind them). Furthermore, Dewan and Shepsle argue that it is important to consider the role of information asymmetry: politicians are better informed about politics than voters, allowing them to conceal certain actions and issues.

Based on the literature review, it can be concluded that there is scope for further research into voter behaviour during electoral campaigns, particularly to understand and model the trajectories of changes in voter preferences as campaign influences take hold and swing support from one candidate or party to another.

Our paper considers electoral campaigns as a stochastic dynamic process of transition between possible system states over time. The change of states (the size of changes in voter preferences in this case) is defined in the model by randomly-arising shifts in voter preferences, together with potential self-organisation within the population (regarded as a system) and the influence of "memory" effects.

In the next section, we calculate the statistical characteristics of shifts in voter preferences (expected value and variance), including their differences when analysed across various time intervals, before describing our new theoretical model and comparing its analytical results with the actual election data observed.

3. Selecting data for analysing social processes and deriving histograms of their statistical distributions

Many different social and socio-technical processes could have been chosen as a case study for analysing and modelling system dynamics. However, creating and verifying the model requires a significant volume of real data. Our search for an appropriately rich information source showed that the largest statistically reliable and publicly available database currently in existence relates to electoral processes, notably the following resource: http://www.realclearpolitics.com/. We therefore chose electoral processes to analyse and develop our model.

A further reason why we chose electoral processes for our modelling is the huge interest in analysing the dynamics of changes in voter preferences during pre-election campaigns and in examining the prediction of results. The events of 2016, notably the Brexit vote and the US elections, revealed weaknesses in the contemporary models used in election predictions. Most research and survey services predicted a victory for Hillary Clinton and predicted that the UK would vote against leaving the EU. So, despite being based on high-quality statistical data, polls and surveys of voter preferences, most forecasts turned out to be wrong. This is partially due to the high volatility of data obtained in polls, but the unexpected outcomes – going beyond what would be expected from random fluctuations and sampling errors – underline the view that existing predictive approaches are based on models that are unsatisfactory. With hindsight, the unexpected outcomes in both cases are seen to be partly related to unorthodox campaign messages and campaign-influencing policies. The models should have signalled the effects of these practices based on fluctuations in earlier data, but they did not.

Fig. 1 below presents the polling data of electoral preferences. The polling was held in the US over 500 days, from 1 July 2015 to 7 November 2016, during the 2016 presidential election campaign (RealClearPolitics, 2016). To avoid any misunderstanding, it is
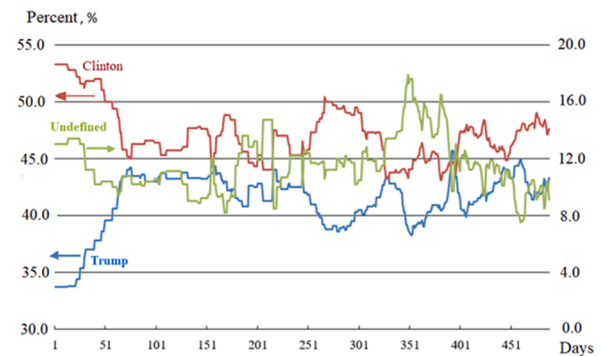

Fig. 1. Electoral preferences of US voters during the 2016 Presidential campaign – Trump, Clinton and uncommitted (undefined) voters.

important to note that the data in Fig. 1 refers to two different scales. Data for Clinton and Trump are presented on the left vertical axis, while data for uncommitted voters are presented on the right vertical axis. The particularities of the US electoral system, where the final decision is made by the Electoral College, are beyond the scope of the present paper. It should be noted, however, that in absolute terms, Clinton received several million more votes than Trump.

Analysis of the data collected is of particular interest both for explaining the outcomes of past elections and for identifying potential opportunities for forecasting future election campaign outcomes, especially when candidates obtain very close results in voter preference polls.

To create and analyse a model using the dynamics of electoral processes, it is necessary to process the observed data and describe the primary variables, which will then lay the foundations for the new model. In this case, the share or percentage of voters favouring a specific candidate (voter preference) is the key characteristic of the process studied.

To process the observed data and determine the probability density functions of the size of voter preference changes, we used the following algorithm:

1) Select all values of recorded voter preferences for a specific time range: day, week, month, etc.
2) Calculate the sizes of changes in voter preferences for various time intervals: day, week, month, etc.
3) The sets of values obtained for each of the intervals measured must be sorted in ascending order (from negative to positive). For each calculation interval, build distribution density histograms for the various values.
4) Calculate distribution moments using histograms for each of the time intervals (average value, expected value, variance, etc.).
5) Develop dependencies – determine the relationship between the distribution moments of voter preference shifts and the various time intervals.

The algorithm allows us to generate a time series characterising the dynamics of voter preferences. The histograms of the probability density of voter preference shifts in the 2016 US presidential election campaign are presented in Fig. 2 for different time intervals. Histograms for undecided voters are not shown in Fig. 2 and subsequent figures, since the total number of voters is 100% and their data are easily calculated if voter preferences for Trump and Clinton are known.

In Fig. 2, the various observed shifts (their probabilities) are indicated on the y-axis, while the x-axis indicates the centres of their specific numerical ranges.
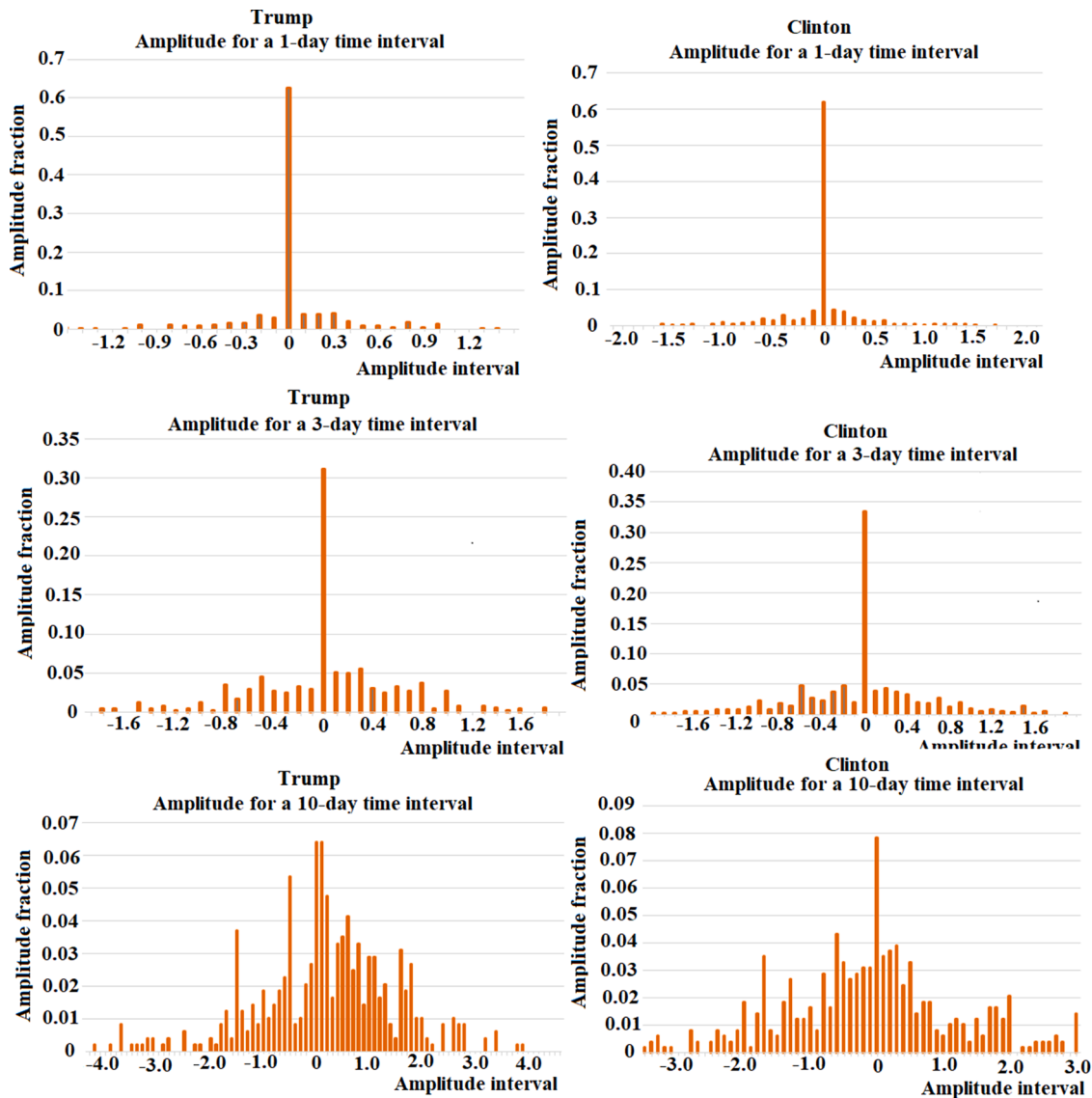
Fig. 2. Histograms of probability density for different shifts in voter preferences in the 2016 US presidential election campaign.

4. An analysis of the histograms of observed statistical distributions

The 2016 US presidential electoral campaign lasted for a total of 500 days. Assuming that the number of the analysed points in the time series in the most minimal case is half of that in the initial series, the dependencies for expected values and variances were calculated from the 1st to the 250th day.

Visual analysis of the data represented in Fig. 2 shows that for small time intervals (one day), histograms of shifts in voter preferences have a big central peak near to the zero value. The probability of the shifts being small in magnitude is about 0.6, while large shifts are only observed with low probability. When the time interval for determining voter preference shifts increases, the central peak decreases, the distribution width increases and the spectrum shows oscillations (this is especially prominent for the amplitude calculated for a time interval of 10 days, see Fig. 2). This indicates significant complexity in the processes observed.

We now consider the behaviour of the statistical parameters of distributions for the processes observed. Fig. 3 provides a dependency chart showing the relationship between the expected value of voter preference shifts and the time intervals for which they were calculated, using the

observed data.

The results obtained (see Fig. 3) indicate that the distribution of voter preference shifts, in function of the time interval observed, differs significantly from the normal distribution. In a normal distribution, the expected value does not depend on the time interval (it is constant or zero). In this case, the processes analysed are thus not entirely random.

We now consider the behaviour of the variance of preference shift values. Fig. 4 presents a dependency chart showing the relationship between the variance of preference shift values and the time interval for which they were calculated, using the observed data.

It should be noted that under a normal distribution, their variance would have a linear dependency on the calculation time interval, however this is not the case (see Fig. 4), meaning that the nature of the distribution differs significantly from the normal rule.

We also note that there is a small degree of asymmetry and oscillation on the amplitude distribution histograms for various time intervals (see Fig. 2), which also indicates that the observed processes are complex and that their statistical characteristics do not correspond to the normal law of distribution. This suggests that self-organisation and memory, which cause non-linearity, may have an effect on the dynamics of the processes.
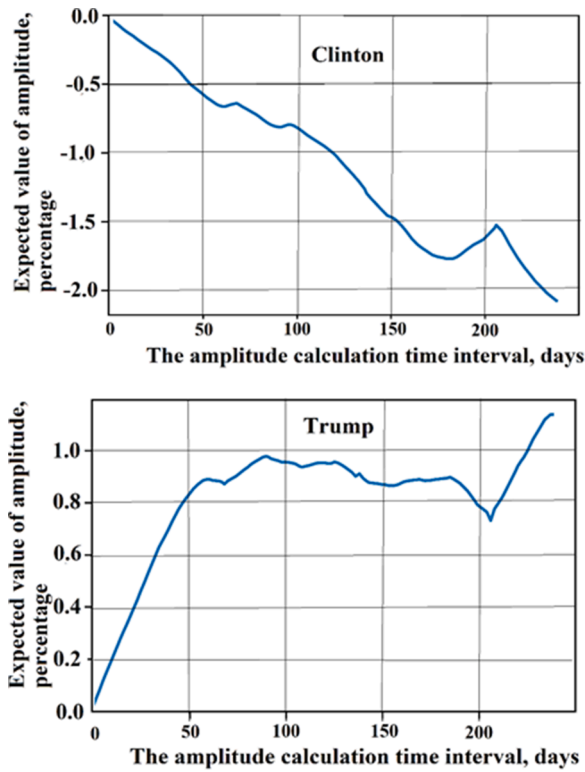
Fig. 3. Dependence of the expected value of the amplitudes of deviations in voter preferences in the 2016 US presidential election campaign.

5. A stochastic dynamics model of forming states in a social system, considering self-organising processes and the potential for memory

By statistically processing the observed data and developing histograms of probability density for voter preference shifts for the time interval for which they were calculated, we were able to develop a theoretical model and compare the results with the observed data.

5.1. Developing probabilistic difference schemas of transition between states to obtain the primary equation of the model

A multitude of possible voter preference shifts for any time interval t (a rather high value) will be denoted as X. We assume that the time interval t comprises smaller parts τ. In this case, any value of time interval t can be represented as $t_h = h \tau$, where h is the step number τ ($h = 0, 1, 2, 3, ...N$). The value of h can be construed as discrete time, the

quantum of which equals τ. The amplitude for the selected time interval t is designated as $x_h : (x_h \varepsilon X)$. The analysis of observed values represented in Fig. 2 shows that x_h can take any positive or negative value. The observation probability of high values is significantly lower than the probability of small values.

We propose that the amplitude x_h, when changing discrete time h per unit, can be increased for a specific small value ε or decreased for a specific value ξ (generally, $\varepsilon \neq \xi$ can be an increasing or decreasing trend). This enables amplitude x_h for the interval of discrete time h to be represented as sets for various ξ and ε (according to the sign). Furthermore, we can assume that neither value x_h can be stationary; for example, it must change when h changes per unit. This is due to the fact that x_h must be changed by ξ or ε when h changes.

The value of parameters ε (we call ε a trend increasing voter loyalty) and ξ (we call ξ a trend decreasing voter loyalty) can be influenced by many different social, economic or other factors. An economic factor such as rising unemployment will reduce voter loyalty; and vice versa, falling unemployment will improve loyalty to the current president. Social factors such as rising crime rates or mounting interethnic tensions connected with increased migration will lower the current president's ratings. Other factors can also be influential. It is probable that the Covid-SARS pandemic had a significant effect on the results of 2020 US presidential elections. Donald Trump denied its danger for a considerable period of time, allowing his opponents to blame him for the death of a significant number of Americans, and this may in turn have changed the impact of criticism in general – making it a more potent element of the mix of factors. Parameters ε and ξ have a complex social nature and cannot be defined by a single factor. They result from the interaction of many different reasons, explicit and implicit. Nonetheless, the progress of a campaign is tracked at a holistic level. Changes in overall voting intention are the basis for assessing the trajectory, and extrapolations based on this trajectory require a model. We therefore propose a model that takes into account parameters ε and ξ in a way that recognises memory and self-organisation.

The current model can be used to analyse the processes observed at the initial stage of a campaign (for example, its first half) and can define the parameters ε and ξ for the further forecasting of results. Employing these details and our probability formulae, our model does not need to represent in detail the nature of all the factors that can affect parameters ε and ξ.

We will now calculate the probability $P(x, h)$ that the amplitude of shifts in voter preferences for a specific interval of discrete time h will be equal to x. We state that:

- $P(x - - \varepsilon, h - - 1)$ is the probability that the amplitude equals $(x - - \varepsilon)$ for some $(h - - 1)$;

- $P(x + \xi, h - - 1)$ is the probability that the amplitude equals $(x + \xi)$ for some $(h - - 1)$;

- $P(x, h - - 1)$ is the probability that the amplitude equals x for some
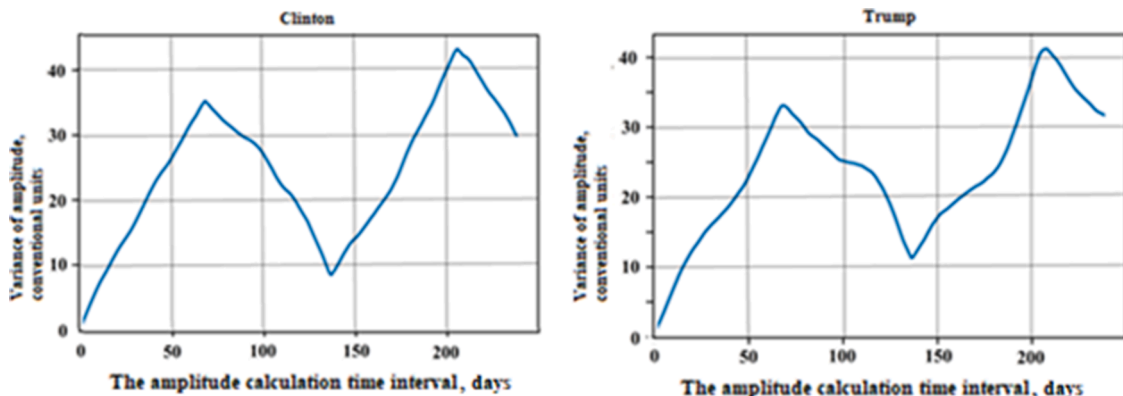

Fig. 4. Dependence of the variance of voter preference shift values in the 2016 US presidential election campaign.

$(h - - 1)$.

The probability $P(x, h)$ that the amplitude of shifts in voter preferences for the interval of discrete time h will be x (see Fig. 5) can be determined as follows:

$$P(x,\ h) = P(x - - \varepsilon,\ h - - 1) + P(x + \xi,\ h - - 1) - - P(x,\ h - - 1) \tag{1}$$

The human factor introduces uncertainty into the effect on processes and creates stochastic behaviour. It also creates opportunities for self-organisation and defines the availability of memory of previous actions. The proposed model must therefore take this into account. In order to consider memory, we define probabilities $P(x - - \varepsilon, h), P(x + \xi, h)$ and $P(x, h)$ based on the states in the previous step $h - - 1$. Diagrams of the respective transitions can be depicted in a similar way to the diagram presented in Fig. 5. Knowing that ε and ξ are constant values, we obtain the following algebraic equation for transition probability:

$$P(x, h+2) = \{P(x - 2\varepsilon) + P(x - \varepsilon + \xi, h) - - P(x - \varepsilon, h)\} + \{P(x + \xi - \varepsilon, h) + + P(x + \xi, h) - P(x + \xi, h)\} - P(x - \varepsilon, h) - P(x + \xi, h - 1) + P(x, h) \tag{2}$$

Given that $t = h \cdot \tau$, where t is the process time, h is the step number and τ is the length of a single step, we move from h to t and then undertake respective dissolutions of the Taylor series to obtain the following:

$$\frac{dP(x,t)}{dt} = a \frac{d^2 P(x,t)}{dx^2} - b \frac{dP(x,t)}{dx} - c \frac{d^2 P(x,t)}{dt^2}, \tag{3}$$

Where $a = \frac{\varepsilon^2 - \varepsilon\xi + \xi^2}{\tau}; b = \frac{\varepsilon - \xi}{\tau}; c = \tau$ Equation factor $\frac{dP(x,t)}{dx}$ describes an organised transition to the another state where voter loyalty is increased ($\varepsilon > \xi$) or decreased ($\varepsilon < \xi$) (voters' activities not caused by random factors). Equation factor $\frac{d^2 P(x,t)}{dx^2}$ describes a random change of state (uncertainty of change). Equation factor $\frac{dP(x,t)}{dt}$ can be defined as the speed of total change of the system state over time. Equation factor $\frac{d^2 P(x,t)}{dt^2}$ describes the process whereby states become sources of other states (self-organisation and acceleration of both organised $\left(\frac{dP(x,t)}{dx}\right)$ and random $\left(\frac{d^2 P(x,t)}{dx^2}\right)$ processes).

As discussed above, the dynamics of social processes (for example, elections) are very complex in nature and are influenced by numerous social and economic factors. In addition, there is always a possibility that other, essentially unpredictable, factors may evolve. For example, it would be possible to hypothesise that the Covid-SARS pandemic had a significant impact on the results of the 2020 US presidential election campaign. The result of a campaign ultimately depends on a combination of factors, many of which are hidden and impossible to observe. It is therefore important to search for new approaches to analyse and model processes in complex social systems. The novelty of our model is seen in
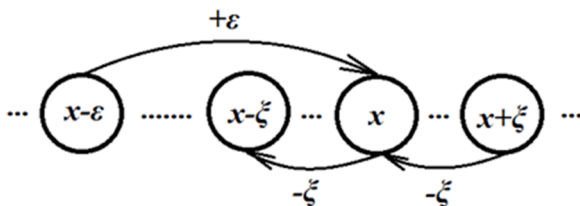
our innovative mathematical approach, which allows us to take into account memory and the potential for self-organisation while describing the processes occurring in complex social systems in a purely statistical manner. The newly developed model thus adds value to the development of the theory of analysing and modelling complex systems. It is important to note that this new model does not contradict pre-existing models. It instead enhances them by considering self-organising processes and the memory of previous states.

In our model, a "shock" appears because of self-organisation and the system's memory of previous states. Such shocks are not shocks caused by economic, social and mass media influences. Our model instead provides a mathematically coherent picture of how shocks may arise in the progression an electorate's views from one level to another – independently of any external influences. However, the approach we are using to describe complex processes also allows us to consider the "shocks" caused by economic, social and mass media factors. In particular, to account for the effect of the external influence of the news on electoral processes, Eq. (3) needs to be enriched by one factor, in order to account for such an influence. This factor can be presented in the equation as follows: $\Sigma k_i \bullet \delta(t-t_i)P(x,t)$, where $\delta(t-t_i)$ are delta functions, taking a value of 1 at times ti when i-events begin to occur (and 0 at other times), and k_i are certain numeric coefficients that are positive when an i-event creates a positive trend and negative otherwise. In the theoretical approaches described above, the appearance of a "shock" is not modelled by a probability distribution; whereas in the new model, a "shock" is considered to be one state of a complex system that can be caused by its potential for self-organisation and memory of previous states.

5.2. Formulating and solving the boundary value problem – finding the distribution function for the amplitude of voter preference shifts

If we consider the function $P(x, t)$ to be continuous, we can go from the probability $P(x, t)$ (Equation 5) to the probability density $\rho(x, t) = \frac{dP(x,t)}{dx}$ and formulate the boundary value problem in order to find the distribution function for various amplitudes of voter preference shifts for an arbitrary time interval t.

Boundary conditions for possible amplitudes of voter preference shifts are selected based on the following assumptions. Our analysis of the statistical data shows that the probabilities of observing high deviation amplitudes (several percent decrease or increase) in voter preferences are extremely small within the time intervals studied. We can therefore assume that the density function of the amplitude probability must rapidly decrease and become zero for high values, meaning that the limit conditions can be set as follows:

$$\rho(x,t)_{x=\infty} = 0 \tag{3a}$$

$$\rho(x,t)_{x=-\infty} = 0 \tag{3b}$$

We set the first initial condition as a delta function based on the fact that for the time interval $t = 0$, the amplitude of $xo = 0$ is only possible when:

$$\rho(x,t)|_{t=0} = \delta(x-0) = \begin{cases} \int \delta(x-0)dx = 1, x = 0 \\ 0,\ x \neq 0 \end{cases} \tag{3c}$$

We must then set the second initial condition: $\frac{\partial \rho(x,t)}{\partial t}|_{t=0}$. This is the condition that sets the change speed of probability density for any value

Fig. 5. Schema of transitions between the amplitudes in function of the change in the value of h per unit (Zhukov et al., 2017; Zhukov et al., 2020).

of amplitude.

During an electoral campaign, there are various types of voters whose behaviour will define some amplitude of change in preferences. Summing up the various types of voters can cause some amplitudes to strengthen while others weaken. This can be influenced by coherence (concordance) in the actions of the campaign candidates. Finally, this may cause periodicity for some amplitude values, e.g., waves. The value of $\frac{\partial \rho(x,t)}{\partial t}$ is the speed of change of function $\rho(x,t)$ and can be stated as follows:

$$\frac{\partial \rho(x,t)}{\partial t} = \lim_{\Delta t \to \tau} \frac{\rho(x + \Delta x, t + \Delta t) - \rho(x,t)}{\Delta t}$$

Given that $\Delta t \to \tau$ (according to the conditions of processes), we set the second initial condition as follows:

$$\frac{\partial \rho(x,t)}{\partial t}\bigg|_{t=0} = \lim_{\Delta t \to \tau} \frac{\rho(x + \Delta x, t + \Delta t) - \rho(x,t)}{\Delta t}\bigg|_{t=0} = \frac{\rho(x + \Delta x, 0 + \tau) - \rho(x,0)}{\tau} = \frac{\rho(x + \Delta x, 0 + \tau) - \delta(x - 0)}{\tau} = \frac{1}{\tau} \cdot \psi(x) \cdot \delta(x - y),$$

where $\psi(x)$ are periodic functions (for example $\psi(x) = \begin{cases} cos\left\{2\pi \frac{x}{\lambda}\right\} \\ sin\left\{2\pi \frac{x}{\lambda}\right\} \end{cases}$)

that can be defined when analysing the observed data and specifying the model. Function $\delta(x - y)$ is related to the fact that the numerator of the fraction contains $\delta(x - 0)$, while also considering that for each point x there is an individual value $\psi(x)$ (y acts as the current value of x).

Solving the boundary problem for Eq. (3) using conditions 3a, 3b, 3c and the periodic second initial condition (for $\frac{\partial \rho(x,t)}{\partial t}\big|_{t=0}$) enables us to obtain the following dependency for the function of the density of voter preference probability deviation amplitude for the time interval:

$$\rho(x,t) = \frac{\tau \cdot \mathscr{U}(t - k) \cdot e^{\frac{(\varepsilon - \xi)x}{2(\varepsilon^2 - \varepsilon\xi + \xi^2)}} \cdot e^{-\frac{t}{2\tau}}}{2\sqrt{\varepsilon^2 - \varepsilon\xi + \xi^2}} \left\{ \frac{1}{\tau}\left\{\frac{1}{2} + \psi(x)\right\} \sum_{n=0}^{\infty} \frac{\left\{\omega^2 \left\{t^2 - k^2\right\}\right\}^n}{4^n \cdot (n!)^2} + \frac{t}{t^2 - k^2} \frac{2n \cdot \left\{\omega^2 \left\{t^2 - k^2\right\}\right\}^n}{4^n \cdot (n!)^2} \right\} \tag{4}$$

where $\omega = \sqrt{\frac{\varepsilon\xi}{4\tau^2(\varepsilon^2 - \varepsilon\xi + \xi^2)}}$; $k = \frac{|x|\tau}{\sqrt{\varepsilon^2 - \varepsilon\xi + \xi^2}}$; $\psi(x) = \begin{cases} cos\left\{2\pi \frac{x}{\lambda}\right\} \\ sin\left\{2\pi \frac{x}{\lambda}\right\} \end{cases}$ – periodic

functions; $\mathscr{U}(t - k)$ - Heaviside step function.

$$u(t - k) = \begin{cases} 0, & \text{if } t < k \\ \frac{1}{2}, & \text{if } t = k \\ 1, & \text{if } t > k \end{cases}$$

Value λ in the periodic function $\psi(x)$ expresses the wavelength of the oscillation process. When there are no oscillations, $\psi(x) = 0$. Wavelength λ is linked to model parameters ε and ξ. Eq. (4) belongs to wave-type equations describing wave processes. Using the properties of the Heaviside function $u(t - k)$ for $t < k$, a conclusion can be drawn whereby $k = \frac{|x|\tau}{\sqrt{\varepsilon^2 - \varepsilon\xi + \xi^2}}$ means that the wave process propagation is delayed to the time k (k is the delay time for the process) ($k = \frac{|x|}{V_0}$, where $V_0 = \frac{\sqrt{\varepsilon^2 - \varepsilon\xi + \xi^2}}{\tau}$ $= \frac{\lambda}{\tau}$ is the speed of wave process propagation; $\lambda = \sqrt{\varepsilon^2 - \varepsilon\xi + \xi^2}$ is the

wave leng

When there are no oscillations, $\psi(x, \lambda) = 0$. This corresponds to the condition where there is zero speed of density change in the probability of any amplitude at the initial point in time.

We note that the normalisation condition is fulfilled for the function of

$$\rho(x,t) : \int_{-\infty}^{-\infty} \rho(x,t)dx = 1.$$

As mentioned earlier, we can account for social and economic factors, which can cause a sharp change in the trend of social process development (a "shock"), by describing the key media events that may influence those factors. First, we need to select the media events published on a date t_i and enhance Eq. (3) using the following factor: $\Sigma k_i \bullet \delta(t - t_i)P(x,t)$.

Next, we need to resolve the corresponding boundary value problem and obtain the corresponding probability density functions of transitioning from one state to another. Finally, the solution obtained should be analysed and matched to the real influence of news events on the processes observed.

5.3. Analysis of the model considering the influence of self-organising processes and memory on the amplitude of vote preference shifts

Fig. 6 presents the modelling results obtained for the dependencies of the amplitude probability density deviations for the time interval considered for various sets of parameters ε and ξ using Eq. (4).

When we increase the wavelength $\lambda = \sqrt{\varepsilon^2 - \varepsilon\xi + \xi^2}$, the number of oscillations on charts of deviation for the amplitude probability density of voter preferences decreases, and when we decrease the wavelength, this number increases. With increasing time intervals, the width and number of oscillations grows, and the distribution height falls.

For $\varepsilon > \xi$, the probability density maximum shifts to the right, and for $\varepsilon < \xi$ it shifts to the left. When the time intervals increase, the distribution height falls and the width and number of oscillations grow. Furthermore, distribution asymmetry is observed in relation to the maximum line.

If $\psi(x, \lambda) = 0$, the oscillations disappear and other distribution behaviour characteristics continue their trend. If we select a different set of parameters ξ, ε, τ and t, the behaviour of amplitude probability densities remains the same, despite the fact that numerical values and positions of maximums on the chart change.

Our analysis of the distribution histograms (see Fig. 2, higher values of time interval considered) shows oscillations and weak symmetry, and the maximum distribution density is slightly shifted from zero when the time interval is increased. Comparing these observations with the results of theoretical modelling (see Fig. 6) indicates that the parameters of processes ε and ξ can have close or equal values and that our model corresponds well with the observed data.
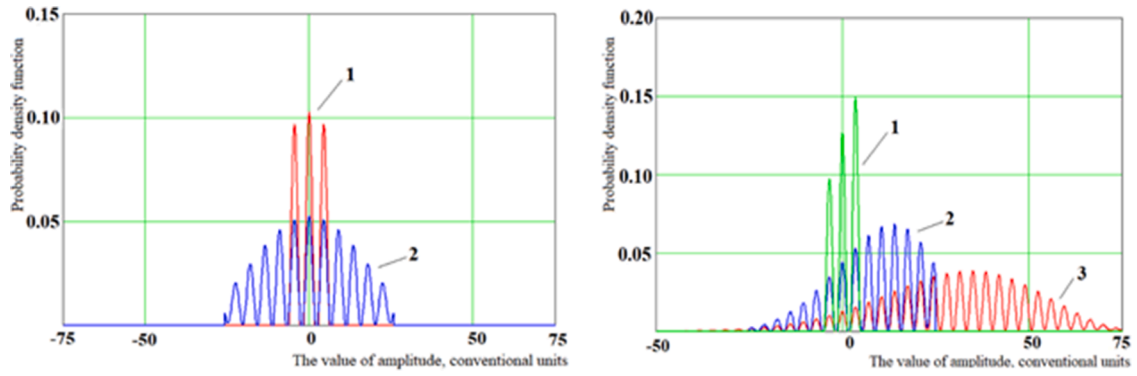
Fig. 6. Dependencies of the amplitude probability densities on the time period of their calculation for the electoral preferences model, taking into account self-organisation, memory and oscillation.
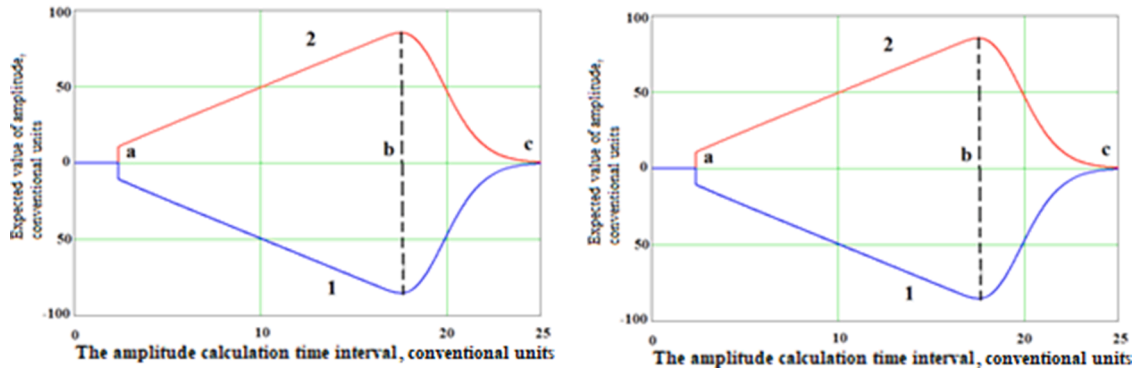

Fig. 7. Dependencies of the amplitude probability densities on the time of their calculation for the electoral preferences model, assuming self-organisation and memory.
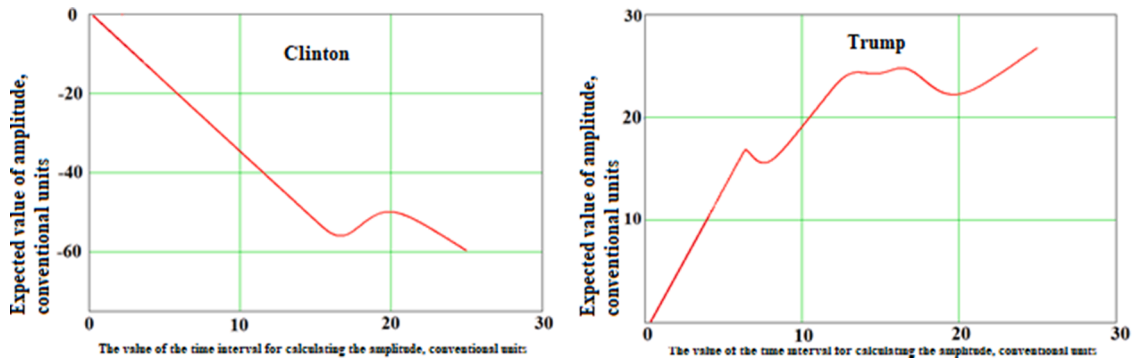

Fig. 8. Results of theoretical modelling of the dependence of the expected value of voter preference shifts (for several processes with different sets of parameters).

Fig. 7 presents the theoretical dependency of the expected amplitude value on the time period considered and on the sets of parameters (ξ and ε).

The average values of deviation amplitude (expected value – $\mu(t)$) and variance – $\sigma^2(t)$ are calculated using expression (4) as follows:

$$\mu(t) = \int_{-\infty}^{+\infty} x \cdot \rho(x,t)dx,$$

$$\sigma^2(t) = \int_{-\infty}^{+\infty} x^2 \rho(x,t)dx.$$

Fig. 7 represents the dependencies of the expected value and variance of amplitude value on the time interval considered, obtained using the present model (for various sets of parameter values ε and ξ). The calculations show that for $\varepsilon > \xi$, a rise in amplitudes is observed, and

their values are within the positive area in Fig. 7 (line no. 2). For $\varepsilon < \xi$, fall amplitudes are observed and their expected value lies within the negative area (line no. 1 in Fig. 7). It should be noted that for sets of parameters with the inversion of ε and ξ: ξ=0.70; ε=0.20 and τ=0.10 and ξ=0.20; ε=0.70 and τ=0.10, the variance behaviour is of the same nature due to the fact that it is a squared quantity with no negative values.

The position of maximums and other process characteristics depend on the selection of model parameter values. When the values of model parameters ε and ξ are increased, the expected value of amplitudes decreases, and the maximum moves to the area of small-time intervals.

The dependencies of the expected value and variance of the observed voter preference shifts on the time of their calculation, obtained from sociological data (see Figs. 3 and 4), differ in type from the modelled results represented in Fig. 7. However, by suggesting several processes to select voter preferences for each candidate with different weight
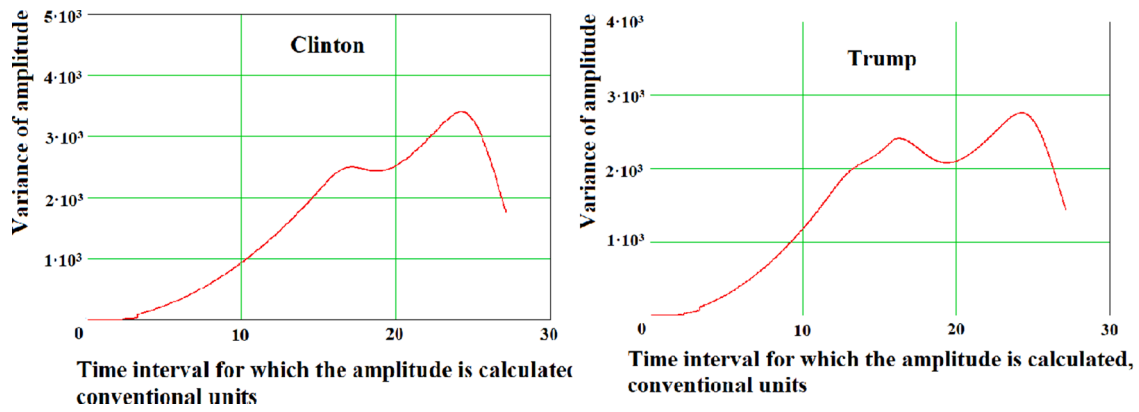
Fig. 9. Results of theoretical modelling of the dependence of voter preference shifts (for several processes with different sets of parameters).

coefficients α_1, α_2.... and several sets of parameters, we obtain the mathematical modelling results presented in Fig. 8, which are a good match for the observed data. We note that the weight coefficients of each process can be different and vary between 0 and 1. The presence of several processes for voting preference dynamics for the same candidate can be caused by the existence of different groups of voters with various types of behaviour when selecting preferences. The values of weight coefficients for the processes discussed above can depend on the number of voters in each group. Such groups and the ratio of their elements (weight coefficients) can be defined through sociological polls.

We consider the changes in voter preferences for Hillary Clinton as the sum of several processes taking place in parallel, each with its own set of parameters ε and ξ. For the I-th process (decrease): $\xi_1 = 0.45$; $\varepsilon_1 = 0.15$; $\tau = 0.1$; $\alpha_1 = 0.80$. For the II-th process (decrease): $\xi_2 = 0.75$; $\varepsilon_2 = 0.20$; $\tau = 0.1$; $\alpha_2 = 0.20$. Similarly, we can represent the changes in voter preferences for Donald Trump as the sum of several parallel processes. For the I-th process (increase): $\xi_1 = 0.15$; $\varepsilon_1 = 0.45$; $\tau = 0.1$; $\alpha_1 = 0.65$; for II-th process (increase): $\xi_2 = 0.20$; $\varepsilon_2 = 0.75$; $\tau = 0.1$; $\alpha_2 = 0.15$; for the III-th process (increase): $\xi_3 = 0.25$; $\varepsilon_3 = 0.95$; $\tau = 0.1$; $\alpha_3 = 0.10$; for the IV-th process (increase): $\xi_4 = 0.30$; $\varepsilon_4 = 1.70$; $\tau = 0.1$; $\alpha_4 = 0.10$. Fig. 8 presents the results of the theoretical modelling of the dependency of the expected value of voter preference shifts on the time interval observed, while Fig. 9 shows variance.

We should note that if $\xi = 0.45$ and $\varepsilon = 0.15$ in one of the processes for one of the candidates (or if ε and ξ have other values), the values $\xi = 0.45$ and $\varepsilon = 0.15$ will be inverted for the other candidate in the same process. This can be explained by the fact (without considering undecided voters) that the decrease in parameter values for one candidate will cause an increase in the parameters for the other (preference selection processes are related).

Comparing our observational data (see Figs. 3 and 4) with the theoretical modelling results (see Figs. 8 and 9) shows that (with respect to modelling roughness) it is possible to select parameter values that produce a model that is highly compliant with the observational data.

It is important to note that when considering two or more parallel processes of voter preference shifts for the same candidate, we can select sets of parameters ξ, ε and τ that will satisfy both the observed histogram of the distribution of voter preference shifts, and the dependencies of their expected value and variance over time. It is also possible to select numerical values for the weight coefficients of each process (0 to 1).

The results obtained enable us to conclude that the model has a general capability for describing the dynamics as they have been observed in electoral processes and that it models characteristics of the systems involved in ways that can be used to provide better descriptions and possible avenues for predicting the results of electoral campaigns and to aid decision-making.

6. An algorithm for forecasting electoral campaign results and decision-making

Our proposed stochastic model of voter preference formation based on self-organising processes, memory and oscillations as described by Eq. (4) is a significant addition to the modelling tools that can be used to predict results and aid decision-making and captures behaviour that is not well described in previous models.

As an example of its utility, consider Fig. 7. Analysis of the model shows that for the dependency of the expected value and variance of shifts during the time interval for which it is calculated, there is a section (designated as |**0a**| in Fig. 7) where they equal zero. The length of the section |**oa**|, the quantity of extremums, their position (point **b**) and the situation of reaching the function of zero expected value (position of point **c**) depend on the values of parameters ξ; ε and τ. In the section |**oa**|, the expected value equals zero, meaning that with short time intervals, the electoral campaign is completely random and the expected quantity of preferences (with respect to the expected value of the shift) will remain the same as in the beginning of the section |**oa**|.

After point **c**, the expected value also equals zero, and for long time intervals the electoral campaign can be also completely random. An example of the processes underpinning a trend of increasing dynamics in preferences is shown by curve 2 in Fig. 7 for cases when growth amplitudes predominate. If fall amplitudes predominate, the average value of amplitudes is in the negative area (curve 1 in Fig. 7). Determining extremum points is very important for the campaign progress and for decision-making, as it enables us to find the values of model parameters and predict campaign results through calculations.

The overall modelling and prediction algorithm comprises the following steps:

1) Using the data observed for a specific time period (for example, the first half of the electoral campaign), we can build histograms describing the dependency of voter preference shifts on the relevant time interval. We identify dependencies of expected value and variance.
2) We use our model to find the values of parameters ξ and ε. For this purpose, we can use Eq. (4) and the observed characteristics of processes (see Figs. 3 and 4).
3) The results obtained are used to calculate the values of voter preferences at the end of the electoral campaign. If the result does not indicate victory, we must influence the process and change its parameters. This can be achieved through the mass media, for example, by changing the value of parameter ε as desired (increasing our own value or decreasing the rival's value).
4) Repeat steps 1–3 to obtain the desired result.

7. Conclusions

As stated above, our paper provides a framework for systematically representing (modelling) the development of levels of voting intention, a social group process. We analysed the dynamics of voter preferences during the 2016 US presidential election campaign and developed probability density diagrams for preference shifts vs. the time intervals observed. We also developed a model of stochastic dynamics of changing voter preferences using the real data. The model we developed takes into account self-organising processes and the presence of memory. It describes the main characteristics of the processes observed: oscillations and changes of distribution height and width when the time interval changes. When developing the model, we took into account probability schemes of transitions between possible states of social systems (the magnitude of voter preferences). We derived a second-order non-linear differential equation, and formulated and solved the boundary task in order to identify the probability density function of voter preference shifts vs. the time intervals observed. The differential equation of the model contains a factor to take into account the self-organising potential and the role of memory. The potential for oscillations is determined by the initial condition.

We trust that in time, and with further development, the model we introduce here could be used to analyse and forecast the evolution of election campaigns. One of the possible scenarios could, for example, be to use the model to analyse the first half of an election campaign and to forecast its final results – given a continuation of the ongoing influences. To do this, the campaign's evolution would be modelled for various parameters. Then, based on the modelling results, the most favourable trajectory of the process could be chosen. The parameter values could be adjusted subsequently in light of further mass-media or electoral data. For instance, if the first half of a campaign appears to be heading in "the right direction" and a decision is needed whether to continue with the tactics used in the first half or to switch to something else, the model may give clues as to the non-linearity of the effects of the ongoing strategy. If the model predicts a rapid percolation effect, then sticking to the existing tactics may be justified even if a linear extrapolation does not lead to the desired goal.

In the present paper, we did not examine the reasons for individual sharp shifts in electoral preferences. Instead, we investigated the trajectory of changes in preference levels as a result of the ongoing pattern of such shifts (regarded as a particular level of influence on voters), relating this to underlying mechanisms of social effects. We were looking to find consistent quantitative patterns and dependencies between the influences on the system and its reaction. Our model shows essential non-linearity between the influence and the reaction, and most importantly, it incorporates the role of self-organising processes and the presence of memory into the mechanisms of electoral shifts.

We examined statistical data on the 2016 US presidential election campaign. We repeat that it is important to note that this dataset was selected as an example, the reason being that the data are publicly available and of high quality.

It is important to note that there are other forecasting approaches that allow forecasting without sociological surveys specific to the event and without time series analysis. The Allan Lichtman "scoring model", which was developed to forecast election results in the US (Lichtman and Keilis-Borok, 1981), has demonstrated good results in predicting US elections. In the period 1984–2020, the model gave only one wrong forecast (the campaign of 2000, George Bush won). However, the model has rather limited applicability because it was developed based on the specific characteristics of the US context, and that context only. The model is based on a set of 13 political and macroeconomic factors: current state of the economy, charisma of candidates, their successes and failures in foreign policy, scandals, existence of a third candidate, absence of social unrest, etc. It was expressly developed for predicting the results of US election campaigns. Changes in the factors identified are clearly correlated with the results of elections in the United States,

but without a link to an explanatory socio-economic mechanism they cannot be considered as universally applicable to other socio-economic systems.

For other social systems, a special set of criteria could no doubt be developed; however, these criteria can change rapidly over time, particularly in societies that are not as stable as the US. The US economy is characterised by stability in the sense that the victory of a certain candidate belonging to a certain political party does not normally provoke significant shocks. By contrast, the emerging markets (which are often the target of political risk assessments) frequently exhibit social systems in which political dynamics are less stable and less "predictable". We expect that our model could be used to work with less stress-resilient social systems, and it therefore presents a more universal instrument for analysing and forecasting election results.

Our paper provides a framework for systematically representing (modelling) the development of levels of voting intention, and we can summarise our concluding remarks as follows:

1. We analysed the dynamics of voter preference shifts during the 2016 US presidential election campaign. The data we obtained were used to build histograms of voter preference shifts calculated over a range of time periods.
2. Our analysis enabled us to develop a model of stochastic dynamics of voter preference shifts that takes self-organising processes and the presence of "memory" into account, and that can accurately describe the characteristics of the processes observed (occurrence of oscillations, changes in distribution height and width when changing the time interval observed, etc.). In function of the parameters of the model and their interrelations, we showed that it is possible to describe various dynamics of voter preference shifts and their statistical characteristics. The model reflects the occurrence of oscillations and changes in distribution height and width, etc. when the time interval considered changes.
3. We considered probabilistic schemes of transitions between possible states when developing the present model of stochastic dynamics of voter preferences. Using this approach, we derived a second-order non-linear differential equation and we formulated and solved the boundary value problem, generating probability density diagrams for voter preference shifts vs. the time intervals observed. The differential equation of the model contains a factor to take account of self-organisation and memory of previous states. The potential for oscillations is determined by the initial condition.
4. Our stochastic model of the variance dynamics of voter preference shifts based on self-organising processes, the presence of memory and oscillations can be developed to predict the results of electoral campaigns and to aid decision-making.

References

Andrianova, E.G., Golovin, S.A., Zykov, S.V., Lesko, S.A., Chukalina, E.R., 2020. Review of modern models and methods of analysis of time series of dynamics of processes in social, economic and socio-technical systems. Russian Technol. J. 8 (4), 7–45. https://doi.org/10.32362/2500-316X-2020-8-4-7-45 (In Russ.).

Airoldi, E.M., Blei, D.M., Fienberg, S.E., Xing, E.P., 2008. Mixed membership stochastic blockmodels. J. Mach. Learn. Res. 9, 1981–2014.

Barrat, A., Barthèlemy, M., Vespignani, A., 2008. Dynamical Processes on Complex Networks. Cambridge University Press, Cambridge.

Bonica, A., Rosenthal, H.B., Rothman, D.J.C., 2014. The political polarization of physicians in the United States: an analysis of campaign contributions to federal elections, 1991 through 2012. JAMA Intern. Med. 174 (8), 1308–1317.

Centola, D., Macy, M, 2007. Complex contagions and the weakness of long ties. Am. J. Sociol. 113, 702–734.

Dewan, T.A., Shepsle, T.A., 2011. Political economy models of elections. Annual Rev. Political Sci. 14, 311–330.

Easley, D., Kleinberg, J., 2010. Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge University Press, Cambridge.

Gleeson, J.P., Cahalane, D.J., 2007. Seed size strongly affects cascades on random networks. Phys. Rev., E 75, 056103.

Gleeson, J.P., 2013. Binary-state dynamics on complex networks: pair approximation and beyond. Phys. Rev. X 3, 021004.

and Granovetter, M., Soong, R., 1983. Threshold models of diffusion and collective behavior. J. Math. Sociol. 9, 165–179.

Gleeson, J.P., 2011. High-accuracy approximation of binary-state dynamics on networks. Phys. Rev. Lett. 107, 068701.

Hay, J., Flynn, D, 2016. How external environment and internal structure change the behavior of discrete systems. Complex Syst. 25 (1), 39–49.

Karsai, M., Iñiguez, G., Kaski, K., Kertész, J., 2014. Complex contagion process in spreading of online innovation. J. R. Soc. Interface 11, 20140694.

Khvatova, T., Zaltsman, A., Zhukov, D, 2017. Information processes in social networks: percolation and stochastic dynamics. CEUR Workshop Proceedings 2nd Int. Scientific Conference "Convergent Cognitive Inf. Technol. 2064, 277–288.

Khvatova, T., Block, M., Zhukov, D., Lesko, S., 2016. How to measure trust: the percolation model applied to intra-organisational knowledge sharing networks. J. Knowl. Manag. 20 (5), 918–935, 10.1108/JKM-11-2015-0464/.

Kocsis, G., Kun, F., 2011. Competition of information channels in the spreading of innovations. Phys. Rev. E 84, 026111.

Li, J., Chen, Z., Qin, T., 2013. Document Using cellular automata to model evolutionary dynamics of social network. IET Conference Publications, 2013(644 CP) 200–205.

Lichtman, A.J., Keilis-Borok, V.I., 1981. Pattern recognition applied to presidential elections in the United States, 1860-1980: role of integral social, economic, and political traits. Proc. Natl. Acad. Sci. U.S.A. 78 (11), 7230–7234. https://doi.org/10.1073/pnas.78.11.7230.

Mundim, P.S., 2012. The press and the vote in the 2002 and 2006 Brazilian presidential campaigns [Imprensa e voto nas eleições presidenciais Brasileiras de 2002 e 2006]. Revista de Sociologia e Politica 20 (41), 123–147.

Nematzadeh, A., Ferrara, E., Flammini, A., Ahn, Y.-.Y., 2014. Optimal network modularity for information diffusion. Phys. Rev. Lett. 113, 088701.

Piedrahita, P., Borge-Holthoefer, J., Moreno, Y., Arenas, A., 2013. Modeling self-sustained activity cascades in socio-technical networks. Europhys. Lett. 104, 48004. https://doi.org/10.1209/0295-5075/104/48004.

Plikynas, D., Raudys, A., Raudys, S., 2015. Agent-based modelling of excitation propagation in social media groups. J. Experimental Theoretical Artificial Intelligence 27 (4), 373–388. https://doi.org/10.1080/0952813X.2014.954631.

Sigov, A., Obukhova, A., Alyoshkin, A., Zhukov, D, 2017. Mathematical solutions of software for predicting group behavior in a social system based on stochastic cellular automata with memory. CEUR Workshop, Proceedings 2nd Int. Sci. Conference "Convergent Cognitive Inf. Technol. 2064, 175–186.

Singh, P., Sreenivasan, S., Szymanski, B.K., Korniss, G., 2013. Threshold-limited spreading in social networks with multiple initiators. Sci. Rep. 3, 2330.

Wang, A., Wu, W., Chen, J, 2014. Social network rumors spread model based on cellular automata. InProceedings - 2014 10th Int. Conference on Mobile Ad-Hoc and Sensor Networks, MSN 2014. https://doi.org/10.1109/MSN.2014.39.

Watts, D.J., 2002. A simple model of global cascades on random networks. Proc. Natl. Acad. Sci. USA 99, 5766–5771.

Wiener, P., 1954. The human use of human beings. New York.

Wiener, P., 1914. The Highest Good. J. Philosophy, Psychol. Sci. Methods, 11 (19), 512–520.

Yağan, O., Gligor, V., 2012. Analysis of complex contagions in random multiplex networks. Phys. Rev. E 86, 036103.

Zhukov, D., Khvatova, T., Zaltcman, A., 2017. Stochastic dynamics of influence expansion in social networks and managing users' transitions from one state to another. In: In Proceedings of the 11th European Conference on Information Systems Management. University of Genoa, Italy, pp. 322–329, 14-15 September 2017.

Zhukov, D., Khvatova, T., Lesko, S., Zaltsman, A., 2018. Managing social networks: applying Percolation theory methodology to understand individuals' attitudes and moods. Technol. Forecast Soc. Change 129, 297–307.

Zhukov, D., Khvatova, T., Millar, C., Zaltcman, A., 2020. Modelling the stochastic dynamics of transitions between states in social systems incorporating self-organization and memory. Technol. Forecast Soc. Change 158, 120134. https://doi.org/10.1016/j.techfore.2020.120134.

Zhukov, D.O., Khvatova, T.Yu., Zaltcman, A.D., 2021. Percolation transitions in social networks with self-organization and memory. Informatika i ee Primeneniya 15 (1), 102–110 (In Russ.).

Professor Dmitry Zhukov, Doctor of Technical Science, PhD in Applied Science, Federal State Budget Educational Institution of Higher Education "MIREA - Russian Technological University" RTU MIREA (Russia). Currently employed as Professor for the Institute of Cybersecurity and Digital Technologies of RTU MIREA. His-current research is focused on IT, knowledge management, percolation theory, networks and mathematical modelling. Other interests include competence modelling and improving the quality of higher education. Dmitry teaches Mathematical Modelling. He has published his research in *Technological Forecasting and Social Change, Journal of Knowledge Management; Informatics and its applications.*

Professor Tatiana Khvatova, Doctor of Science (HDR) in the field of Management, PhD in Applied Sciences. Currently employed as Professor in Innovation at EMLyon business school (Lyon, France) and as a post-doc at Lappeenranta-Lahti University of Technology (Finland). Her current research is focused on innovation and knowledge management, digital business models, innovation policies and innovation systems. Tatiana teaches Mathematical Methods and Innovation Management. She has published her research in *Technological Forecasting and Social Change, The International Journal of Human Resource Management,* and *Journal of Knowledge Management.*

Professor Carla Millar, PhD in International Business & Management, Professor of International Management & Marketing, University of Twente, NL. Previously Dean of TSM Business School, CEO MDC, City University, London, UK. Her research focuses on leadership, inclusion, innovation, risk and resilience. Carla recently guest edited more than a dozen SIs: Leadership in Global Knowledge Intensive Firms, *International Journal of Human Resource Management,* 2017; Disruptive Technology and Innovation in Society, *Technological Forecasting and Social Change, an international journal,* 2018; Management Innovation in an Uncertain World, *California Management Review,* 2018.

Associate Professor Elena Andrianova, PhD in Engineering, Federal State Budget Educational Institution of Higher Education "MIREA - Russian Technological University" RTU MIREA (Russia). Currently employed as Department Head of ERP for the IT Institute of RTU MIREA. Her current research is focused on IT, ERP, knowledge management and mathematical modelling. Other interests include competence modelling and improving the quality of higher education. Elena teaches the architecture design of information systems.