

An Approach to Analyzing the Entries of HPC Systems in Supercomputer Ratings as an Extension of the Functionality of the Top50 Russia Rating

Vitalii A. Kuleshov¹ , **Dmitry A. Nikitenko¹** 

© The Authors 2026. This paper is published with open access at SuperFri.org

Supercomputer ratings such as Top500, HPCG, Graph500, IO500, and others reflect technology development trends such as performance, energy efficiency, and so on, and serve as a kind of “honor roll” for the global HPC community. Until now, there has been no convenient tool to track the dynamics of domestic systems’ participation over time: when they appeared, how long they remained in the ratings, and how their positions have changed. This work fills this gap. To understand how HPC infrastructure is developing, we need to look not only at individual points reflecting the characteristics and position of systems at a given point in time, but also at their trajectories over time: when and why they appear in the ratings, how long they hold their positions, in which ratings they show greater stability, and in which they disappear. This new capability is offered by the proposed approach and implementation based on the Top50 Russian supercomputer rating: it combines data from all major global ratings for analysis. The solution is built as an analytical extension of the Top50 rating functionality. Visualizations are formed from simple charts of entry history to advanced tools using candlestick charts and transition matrices.

Keywords: hpc ratings, lifelines of supercomputers, top50 supercomputers of Russia supercomputers, evolution of computing systems, high-performance computing systems.

Introduction

The advancement of High-Performance Computing (HPC) remains a key driver of progress in fundamental scientific research, engineering applications, and industrial computing. Modern supercomputer platforms are employed in climate modeling, computational chemistry, deep neural network training, and numerous other tasks requiring scalable computational power. Architectural features and measured performance metrics directly influence the capability to solve new classes of applied problems and the competitiveness of national scientific and technological ecosystems.

Public supercomputer rankings, primarily the Top500 [18], serve not only as a catalog of current achievements but also as a mechanism for revealing and monitoring technological trends. They enable tracking of performance growth dynamics, the impact of energy efficiency, the proliferation of architectures with varying interconnect topologies, and more. Supplementary lists and benchmarks, such as Green500, IO500, Graph500, HPCG, MLPERF, and others, assess specialized system aspects, including energy efficiency, input/output operations, graph processing, and various workload classes. Collectively, they provide a comprehensive overview of the state of the global high-end computing resource pool.

However, each ranking release represents merely a snapshot: for each system, only the position and a set of specialized metrics are recorded at the time of submission. This approach fails to provide a complete picture of the “lifecycle” of physical systems. It does not reflect configuration history (upgrades, partial updates), overlooks naming ambiguities across releases, and hinders the assessment of positional stability over time. Consequently, managerial and research conclusions based solely on individual releases may be incomplete or misleading.

¹Research Computing Center of Lomonosov Moscow State University, Moscow, Russian Federation

For monitoring and analyzing national representation in global rankings, a new approach is required that operates not on individual release data points, but on the trajectories of emergence and evolution of each physical system. Key questions include: when a system first enters the ranking and the duration of its presence; what configuration changes accompanied shifts in position; in which rankings the system demonstrates greater stability; and whether there are “traps” of rapid obsolescence for certain equipment categories. Answering these questions requires data formalization, validation procedures, and a set of normalized metrics enabling correct inter-release comparison.

Supercomputer rankings, and particularly the Top50 list, serve as fundamental indicators of the high-performance computing (HPC) field’s development, enabling the tracking of both regional computational capacity progress and the evolution of architectural design solutions [5, 8]. This study proposes a comprehensive approach to analyzing the inclusion of systems from the national Top50 ranking, developed by the Research Computing Center of Lomonosov Moscow State University (RCC MSU), into international lists. The focus is placed on the Top500 as the primary data source, as it features the widest representation of domestic supercomputer complexes. Within this work, an analysis of the current state and structure of the national Top50 [19] supercomputer ranking website was conducted. Based on this analysis, a data model was developed for tracking the inclusion of Russian systems in international rankings, accounting for their performance characteristics. To ensure correct identification of physical systems and their configurations, an administrative panel was implemented for manual validation and entry of ranking inclusion data. This approach allows explicitly recording which specific system version and configuration is presented in a particular release.

Based on the formed data structure, temporal trajectories (“lifelines”) were constructed for each system mentioned in the rankings, reflecting the history of its presence in international lists. To evaluate not only positional rank but also system quality, two normalized metrics were introduced: the share of achieved performance on the LINPACK benchmark relative to the total performance of ranking participants, and the share of theoretically achievable peak performance relative to the total theoretical peak.

Additionally, a statistical section was implemented dedicated to analyzing system entries into the Top50 and Top500 rankings. This includes an assessment of the system pool refresh rate and visualization of position and performance change trajectories. For extended analysis, 3D representations of lifelines and candlestick charts for aggregated display of metric changes were applied, as well as probabilistic models of transitions between rank categories, constructed based on Markov matrices. The latter are used to diagnose system position stability and identify categories with a higher probability of degradation and rapid obsolescence.

The subsequent sections of the paper develop the outlined ideas: the Literature Review (State of the Art) section examines existing rankings, metrics, and approaches to rank dynamics analysis; the Methodology section formalizes the data model and validation procedures; the Implementation section describes the software solution architecture and implemented visualization components; the Approbation section demonstrates the results of applying the approach to historical Top500 data and the analysis of Russian systems. The Conclusion summarizes the findings and outlines directions for future development.

1. State of the Art

1.1. Global Landscape and Methodological Foundation

The history of systematic ranking of high-performance systems began in 1993 with the emergence of the Top500 ranking, which over three decades has transformed from a simple “honor roll” into the most important tool for technological progress monitoring [11]. The fundamental performance metric in the Top500 is the LINPACK (HPL) test, which measures the speed of solving a dense system of linear equations [4]. Despite criticism for its one-sidedness, the project’s editors emphasize its success as a stable metric for revealing long-term trends, such as the exponential growth of performance and the shift from homogeneous systems to complex hybrid architectures [15].

As the industry developed, the need arose to diversify evaluation criteria. In 2013 the Green500 [13] ranking, focused on energy efficiency (GFlops/W), received official status, which was a response to the environmental crisis and the rising cost of system ownership [1]. At the same time, the Graph500 [12] ranking for assessing graph processing and IO500 [14] for analyzing I/O subsystems appeared. However, as researchers note, the modern landscape remains fragmented: system data is often scattered, and rules for correlating positions across different lists are absent [1, 9].

As the longest-standing regional ranking, the Top50 list has undergone substantial modernization (vivification), adapting to innovations in supercomputer system design and introducing novel approaches to performance comparison across the CIS region [6].

1.2. Architectural Trends and the Problem of System Analysis

An analysis of the 27-year history of the Top500 conducted in 2021 [4] revealed an alarming trend: the HPC community tends to maximize FLOPS (processor performance) at the expense of the balance of other subsystems. The study highlighted the following points:

- Performance Efficiency ($R_{\max}/R_{\text{peak}}$): The average efficiency is only 0.67, and it tends to decrease as architectures become more complex.
- Heterogeneity: Since 2011, the number of systems with accelerators (GPU) has been steadily increasing, and today they provide on average 84% of the peak performance (R_{peak}) of leading supercomputers.
- Role of the Operating System: The transition to open-source software had a huge impact on the growth of computing power. Since the mid-2000s, Linux has become the dominant platform, providing flexibility for experiments with innovative system designs.

An important aspect of the analysis was the proposal of hierarchical models for describing architectures. Nikitenko D.A. [9] points out that standard descriptions in rankings often have a marketing character and lack details about the structure of nodes and interconnections. The model developed by him makes it possible to link the achieved performance with specific architectural features, which is critically important for the scientific comparison of systems from the Top50 and Top500 rankings.

1.3. Dynamics of the Life Cycle and the “Lifecycle”

Researchers are particularly interested not in the static position of a system in the list, but in its evolution over time. D.G. Feitelson [3], in his works on interpreting Top500 data, identified

an empirical regularity: the rank of a given machine on average doubles every year. This means that a system is rapidly “washed out” of the list by new entrants.

Specialized tools have been developed for the visual investigation of these processes. One of the best known is the Top500Vis project by C. Steed [10], which provides interactive capabilities for analyzing distributions and characteristics of systems within individual releases. Figure 1 shows the rankings of countries in the TOP500, but the image is difficult to analyze, and there is no specific information about the systems installed in Russia. However, despite the effectiveness of such tools for analyzing statistical snapshots, methods in the scientific literature that focus on continuous trajectories “lifelines” of specific physical machines are still insufficiently developed. The “lifeline” approach proposed in this article enables the visualization of system ranking history by capturing the temporal intervals of their presence across lists. While these lines are fundamentally constructed from the fact of a system’s inclusion, their interpretation is inextricably linked to research on system and component upgradability. According to recent 2025 data, it is precisely the capability for node-level modernization such as upgrading processors or interconnects that often determines a system’s “lifeline” duration, allowing it to maintain competitiveness amid the rapid technological refresh of the installed base [2].

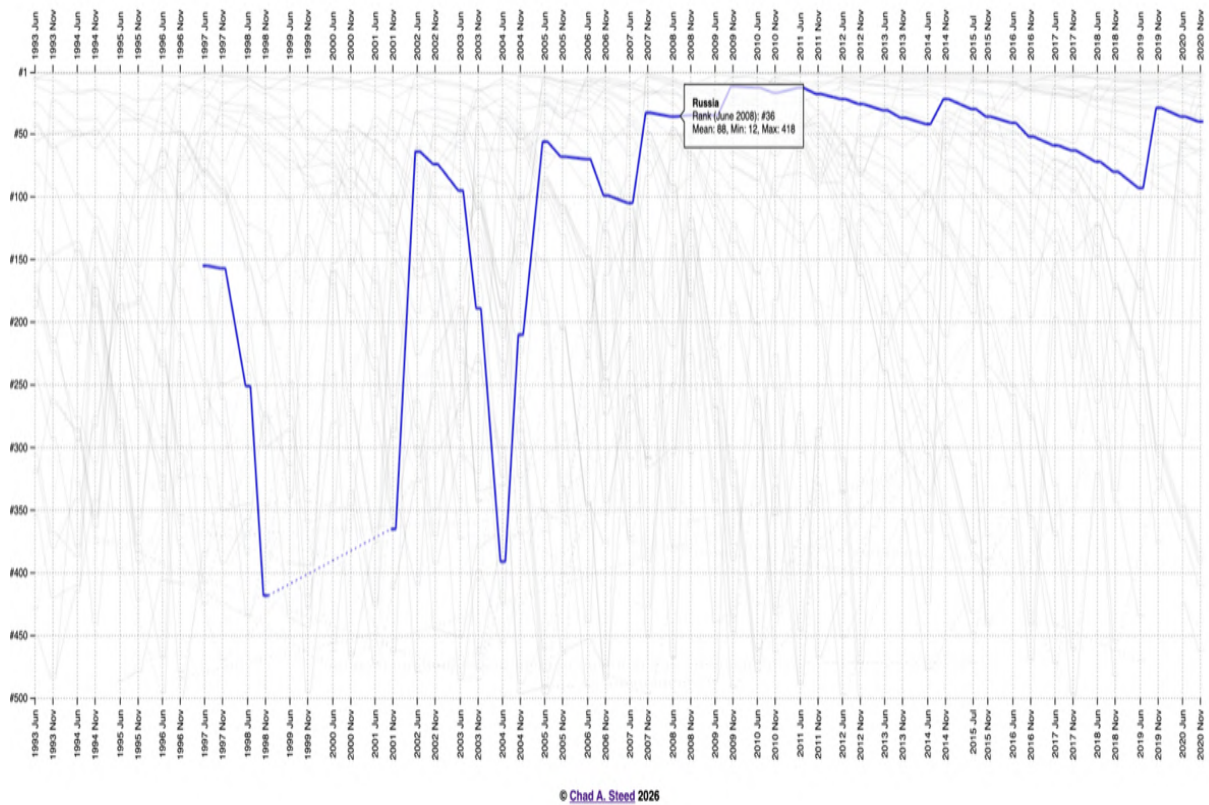


Figure 1. Russia’s portrayal in the Top500Vis project (C. Steed et al.) [10]

Recent studies by Abdessalam Benhari [1] confirm a high system pool turnover rate: 80% of systems remain listed in the Top500 for fewer than 2 years, with an average presence duration of merely 1.4 years, Fig. 2. This dynamics necessitates novel visualization approaches capable of tracking individual physical system trajectories across numerous releases while accounting for hardware upgrades and configuration changes the primary focus of this paper.

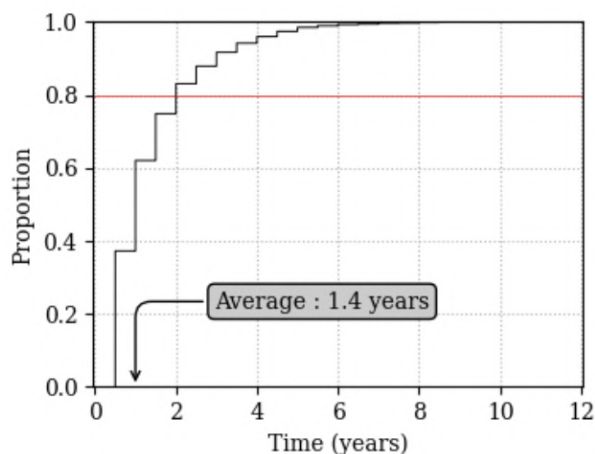


Figure 2. Lifespan of HPC systems in the Top500 (in years) from 1993.
The red horizontal line means that 80% of the systems stay less than 2 years

1.4. The Russian Context and the Shortage of Analytical Tools

In Russia, HPC resource monitoring has been conducted within the framework of the Top50 ranking since 2004, which is the most comprehensive source of data on national systems. However, a methodological gap exists: while global rankings record individual Russian systems, they do so in a fragmented manner, without unified rules for comparison. Current methods for analyzing Russian representation have a number of limitations:

- Publications of the rankings record data at the moment of submission and do not reflect the history of system upgrades.
- Problem of scale: The contribution of a single Russian system to the global computing power is statistically small, which requires the introduction of indicators relative to the national ranking in order to increase the sensitivity of the analysis.
- Lack of end-to-end integration: There is no tool capable of integrating data from Top50, Top500, Graph500, and other lists into a unified development trajectory.

It is important to note here that there are a number of HPC centers in Russia [21], which are potentially more interested in tools, demonstrating the current position and the trends of possessed machines.

1.5. Conclusions and Transition to the Proposed Methods

The literature review shows that despite the abundance of statistical data on supercomputers, the HPC community experiences an acute shortage of tools for in-depth analysis. Existing approaches are either too generalized (as in global rankings) or limited by regional frameworks.

For high-quality monitoring of national representation in the global ecosystem, it is necessary to move from the analysis of “release points” to the analysis of life-cycle trajectories. This requires the creation of a formalized data model, mechanisms for expert validation, and the implementation of advanced visualization methods (such as candlestick charts and transition matrices) capable of diagnosing the stability of positions and scenarios of equipment obsolescence. It is precisely these tasks that the method of analytical extension of the Top50 platform proposed in this work aims to address.

2. Proposed Method for Collecting and Analyzing Data on the Inclusion of Russian Supercomputer Systems in Global Rankings

The modern landscape of global supercomputer rankings represents a fragmented ecosystem in which national and international evaluation systems operate largely autonomously. The Russian domestic ranking Top50 serves as the most complete and detailed source of information on national computing systems, providing diverse data on the configurations, performance, and application domains of domestic supercomputers. At the same time, global lists, including Top500 and a number of other international rankings, have historically recorded individual Russian systems, but they do so in a scattered manner and without unified rules for comparing and analyzing the dynamics of their positions.

At present, there is no reproducible tool capable of integrating heterogeneous sources, harmonizing names and metadata, and then systematically tracking the inclusion of Russian systems in global rankings, constructing time series, and helping to identify stable trends. As a consequence of this fragmentation, there is a shortage of objective information about the international visibility and competitiveness of domestic developments.

The method described below is aimed at addressing this gap: it outlines approaches to the aggregation, normalization, and analysis of data on the inclusion of domestic supercomputer systems in global rankings, as well as tools for the quantitative assessment of their representation and dynamics.

2.1. Conceptual Data Model

The data model for describing inclusions extends the existing data model of the Top50 ranking [6]. The method is based on a conceptual model in which the following key entities and semantic relationships between them are identified:

- **Object**, a logical unit of the hardware/component structure (component, compute node, chassis, supercomputer as an aggregate).
- **Machine**, a record representing a specific computing system in the database. It refers to **Object** via a unique identifier.
- **Relation**, the semantics of relationships between objects. For the described method, the relationship type “precedes” is considered, which is used for versioning and end-to-end tracking of a system across ranking editions.

The extended data model, in turn, relies on the following entities:

- **Extratings_List**, an abstraction representing a specific international supercomputer ranking.
- **Extratings_Editions**, a specific edition of a ranking: issue number, publication date, and sequential release number.
- **Extratings_Entry**, the actual inclusion of a system in a specific edition. It includes the position and a reference to performance indicators.
- **Extratings_Units**, **Extratings_List_Units**, **Extratings_Score**, entities linking measurement units, priorities of units within a ranking, and specific values of performance indicators for a recorded inclusion of a system.

For each entity, mandatory attributes, data types, and integrity constraints are defined to ensure the correctness and consistency of the stored information. The constraints include

primary and foreign keys, uniqueness rules, as well as checks of permissible value ranges for quantitative indicators.

The conceptual data model was designed leveraging expertise from the Algo500 project [5], which established a unified repository of system architectural descriptions. This foundation enables the prospective development of flexible, user-defined rankings based on extensible entities.

2.2. Process of Data Collection and Validation

Manual input makes it possible to minimize the risk of erroneous aggregation and ensures quality control at each stage of observation formation. The function of data validation is assigned to the administrator of the Top50 ranking. This solution is practically justified both from the point of view of the workload and from the point of view of the methodological complexity of automation.

First, international rankings are published no more than twice a year, and the administrator, possessing expert knowledge about the positions of Russian systems in global lists, can efficiently perform verification with minimal labor costs.

Second, automated parsing is complicated by the fundamental heterogeneity of the rankings themselves: they are based on different tests and evaluation criteria. For example, the Top500 ranking is based on the standardized LINPACK benchmark, which ensures comparability of results. At the same time, the Graph500 and IO500 rankings use composite benchmarks that include both regulated components and parts that allow free configuration for a specific architecture in order to maximize performance. The Green500 ranking, in turn, does not impose strict requirements on the size of the problem being solved.

Such variability of evaluation methodologies, along with differences in system names, transliteration, and publication formats, makes automatic extraction and correct interpretation of data unreliable without expert verification.

2.3. Analytical Methods

Within the framework of the proposed method, the dynamics of the presence of Russian supercomputer systems in global rankings is studied. The analysis is aimed at identifying patterns in the change of their positions over time and at assessing the relative contribution of Russian systems to the total computing power of the corresponding national ranking datasets.

The method is based on the use of relative performance indicators in combination with tools of temporal and statistical analysis, which makes it possible to ensure comparability of results between different releases of rankings when their composition and scale change.

2.3.1. Lifelines of the system

One of the analysis methods is the construction of a sequence of inclusions of Russian system's in releases of global rankings. The construction of a life trajectory makes it possible to study the duration of a systems presence in an international ranking list, the dynamics of its positions, and changes in performance over time, taking into account technical upgrades.

2.3.2. Share of achieved productivity and theoretical peak value

Within the framework of the method, an indicator of the share of achieved and theoretically achievable performance is introduced. The share of achieved performance is defined as the ratio of the measured performance of a Russian system in the LINPACK test to the total performance of all systems in the national Top50 ranking. The share of the theoretical peak performance value is calculated in a similar way.

The choice of the national ranking as the normalization base is deliberate. The contribution of a single Russian system to the total computing power of the global ranking is statistically insignificant and does not allow meaningful patterns to be identified. At the same time, normalization relative to the national ranking provides higher sensitivity of the indicator and makes it possible to detect trends that remain invisible when analyzed at the global scale.

The proposed indicators reflect the relative contribution of a system to the total computing power and make it possible to analyze its relative hardware potential in the context of the distribution of computing resources.

2.3.3. Japanese candlesticks

To analyze the dynamics of the positions of Russian systems between releases of global rankings, an adapted representation in the format of “Japanese candlesticks” is used. Within a single release, the best and the worst positions of the represented Russian systems are recorded, which makes it possible to evaluate the amplitude of position changes and identify periods of significant shifts.

Such a representation provides a clear visualization of the direction of changes (improvement or deterioration of position), as well as their degree. The analysis of the sequence of “candles” in dynamics makes it possible to identify periods of relative stability, gradual decline of positions, or sharp changes caused by updates in the composition of the ranking or changes in the competitive environment.

2.3.4. Transition matrices

To formalize the dynamics, a transition matrix is used in which the movements of Russian systems between predefined position categories (1–10, 11–20, 21–30, 31–40, 41–50, 51–100, 101–200, 201–300, 301–400, 401–500) are recorded across consecutive releases of the ranking (quarters). For each system from the sample (for example, Top50 systems included in Top500, although the method is applicable to any ranking), the categories of its positions in adjacent periods are tracked. For each pair of categories $A \rightarrow B$, the empirical transition probability is calculated as the ratio of the number of recorded transitions from A to B to the total number of transitions from A .

Conceptually, this matrix represents a variation of the classical transition matrix of a Markov chain, constructed on the basis of frequency estimates of transitions between discrete states. Its application makes it possible to quantitatively assess the stability of the positions of Russian systems and to identify typical scenarios of their movement in the international ranking space.

Thus, the proposed method provides a comprehensive analysis of the dynamics of the presence of Russian supercomputer systems in global rankings through a combination of relative performance indicators and tools of temporal analysis. The use of shares total computing power makes it possible to take into account the changes in the scale and composition of rankings, while

the application of trajectory analysis, visual representations, and transition matrices makes it possible to formalize the nature and stability of the changes in system positions over time.

The integration of the specified elements within a unified methodological approach ensures the comparability of results between ranking releases and creates a foundation for identifying stable patterns of changes in the position of Russian systems in the global supercomputer space.

3. Implementation of the Proposed Method

The proposed method has been implemented as a server-side application written in Ruby [17], using the PostgreSQL [16] relational DBMS. The Top50 rating project is built on the Octoshell [7] system engine, developed at the Research Computing Center of Lomonosov Moscow State University (RCC MSU) and used for the management and support of computing infrastructure. The use of this engine allowed for the reuse of an already tested data storage architecture, user management mechanisms, and basic service components.

Within the framework of the present work, the system functionality was extended with modules for aggregation, normalization, and analysis of data on the inclusion of Russian supercomputer systems in global rankings. Additionally, an administrative panel for validation and confirmation of data was implemented, as well as a specialized statistics section providing visualization and analytical representation of results.

3.1. Technology stack

- **Programming language:** Ruby on Rails. Used for server logic, aggregation scripts, and business logic. Database extensions were implemented using Ruby migrations.
- **Data storage:** PostgreSQL, serving as the primary database.
- **Data schema management:** Ruby-based migrations, providing versioning of the database structure and reproducibility of deployment.
- **Visualization:** JavaScript and TSX (TypeScript + JSX) for implementing the user interface and interactive visualization components.

3.2. Data Model Implementation

The physical implementation of the data model is carried out in the PostgreSQL DBMS. The table structure corresponds to the conceptual model described in the Methodology section and includes the entities `Extratings_List`, `Extratings_Editions`, `Extratings_Entry`, `Extratings_Units`, `Extratings_Score` and `Extratings_List_Units`.

The database schema definition and its modifications are managed using Ruby-based migrations. This approach ensures reproducibility of the storage structure, schema version control, and consistency of changes when extending functionality.

At the DBMS level, mechanisms are implemented to maintain referential integrity and correctness of relationships between entities, which allows building complex aggregated queries and generating temporal selections without violating data consistency.

The semantics of measurement units are realized through two interrelated entities. `Extratings_Units` serves as a dictionary of units, while `Extratings_List_Units` acts as a junction table, ensuring unambiguous mapping of specific units to a particular rating and recording their priority through the `priority` field. This scheme is necessary because a single rating may

publish multiple performance metrics, and their contribution to the ranking is not always unambiguous. The `priority` field explicitly defines the preference of units during aggregation. Unique correspondence is ensured through the `extratings_list_id` and `extratings_unit_id` fields, which also allows reusing common units across multiple ratings without duplicating records. The `ExtratingsList` entity formalizes the rating itself. The relationship between a specific edition and the rating is established via the `extratings_list_id` field in `ExtratingsEditions`.

3.3. Implementation of the Administration Panel

The administration panel provides input of new entries regarding the inclusion of systems in international rankings. The main workflow involves the following steps:

- First, the administrator selects a system from the Top50 rating directory. Addition of new systems to the directory is handled within the core functionality of the rating: when a previously unregistered system appears, it is first registered in the project database and then becomes available for selection from the dropdown list.
- Second, the administrator selects the rating in which the entry should be recorded. For convenience, the interface automatically populates the fields for performance metrics in the units used by the selected rating. If necessary, measurement units can be changed via a dropdown menu to the units in which the benchmark result was recorded. An integrated conversion mechanism ensures that all values are brought to a unified scale.

The interface also allows extending the directory of existing ratings directly during operation. From the entry creation window, the administrator can create a new rating, specifying its characteristics and selecting the units used for system ranking. Similarly, a function is implemented for creating new measurement units with a link to a specific rating and setting their priority when determining the final system order.

3.4. Implementation of the Lifelines of the Systems

On the page of each system that has been recorded at least once in international rankings, a chart is implemented displaying its lifelines across all considered rankings. By default, all available lists are visualized; if necessary, the user can restrict the display to only the trajectory in a selected rating. Each point on the chart corresponds to a specific entry of the system in a rating edition; hovering over a point shows a tooltip with information about the position, edition number and release date, as well as performance metrics recorded for that entry. Below the chart, a tabular summary of all entries is provided: the table lists the publication date, edition, position, and key metrics – the same information as in the tooltip, but presented in a static and organized form for ease of verification and export.

The section containing aggregated representations is called “Statistics of system entries in Top50 and Top500”. It reflects the current concentration of domestic data, while the structure of the section is designed so that, with future expansion of the dataset, its name or functionality can be modified without changing the storage or display logic of the modules.

The statistics section implements an aggregated display of lifelines for all systems over the entire considered period. In this view, each system is shown as a separate trajectory, which allows visual assessment of the density of position distribution, periods of strengthening or weakening of Russian systems presence in international rankings, and the nature of their simultaneous movement across different ranges of the ranking space. Additionally, a mode is provided for

simultaneous visualization of trajectories in an international ranking and in the Top50 rating, allowing comparison of the systems position dynamics in global and regional contexts. This combined view facilitates identification of consistency or divergence of trends between different ranking datasets.

3.5. Implementation of the Fractions of the Achieved and Theoretical Peak Value

The fractions of achieved and theoretical peak performance are displayed on the page of a specific machine as bar charts. The user can view the entire available time interval to assess the dynamics of metrics; bars are color-coded, and hovering displays a tooltip explaining the corresponding column (release date, metric value, etc.). Joint visualization of the fraction of achieved performance and the fraction of the theoretical peak allows a clear assessment of the divergence between practically realized and theoretically possible performance values and serves as a basis for conclusions about the practical efficiency of the systems software-hardware implementation.

3.6. Implementation of the “Japanese Candlesticks” Metric

The “Japanese candlesticks” visualization is implemented as an interactive client-side component (JS / TSX) within the statistics section. By default, the visual block is displayed in the same view as the lifelines, covering the entire available observation period. On the server side, time series are generated for each ranking publication: the boundary values of the position are computed and used as the “open” and “close” parameters of the candlestick. The resulting series are transmitted to the interface in a format suitable for direct rendering.

The visualization interface supports interactive controls, including selection of the time window (observation period) directly within the chart area, filtering by ranking, and adjustment of the aggregation step. Hovering over a candlestick element displays a tooltip with information about the best and worst position within the release, the edition number and publication date, and the corresponding performance metrics. This design allows the user to obtain both an overview for the entire period and a focused inspection of a specific time interval without navigating to other pages.

3.7. Implementation of Transition Matrices

The interactive implementation allows users to inspect numerical values and percentage estimates for specific category pairs and, if necessary, to restrict calculations to a defined time window or subset (e.g., only Russian systems or only Top50). The transition matrix serves as a tool for quantitatively assessing the stability of positions and identifying typical patterns of movement within the ranking space. The use of the transition matrix is meaningful within a well-defined temporal interval. While it is formally possible to compute the share of transitions from an arbitrary category A to category B relative to all transitions from A over the entire existence of the ranking, the substantive interpretation of such a value may be complicated by changes in the composition of participants and ranking conditions over time. To reduce this ambiguity, the interface by default displays transition probabilities calculated over the most recent two years of publications. This interval provides estimates that reflect current dynamics.

Additionally, the interface offers the option to switch to precomputed values over the last five years of publications, as well as a tool to select an arbitrary time interval within which

transition probabilities will be computed. This functionality allows users to define the temporal scale of analysis according to the specific research task.

The interface includes tabbed display modes: the matrix can be presented either in relative terms (probabilities of transition from category A to category B) or in absolute terms (number of recorded transitions). Although matrices over limited time windows are often sparse, such estimates are more representative and provide a more accurate reflection of the observed dynamics of system positions. Thus, the implemented software tool provides a complete workflow for ranking data: from entry and verification of observations to their analytical and visual interpretation. A centralized data storage model has been implemented, supporting the recording of positions, performance metrics, and their historical evolution, as well as mechanisms for administering reference directories and units of measurement. The interface provides a set of analytical views designed for the study of system dynamics: lifelines, aggregated trajectories, relative measures of achieved and peak performance, transition matrices with configurable time windows, and interactive 3D visualization. Implemented filtering mechanisms, time window selection, and display mode switching (absolute values versus probabilities) ensure analytical flexibility and allow the derivation of interpretable quantitative estimates that reflect the observed dynamics of ranking positions and system performance. Collectively, these components constitute an integrated environment for analyzing the evolution of computational systems over time, ensuring data consistency, reproducibility of calculations, and clarity of the resulting insights.

4. Approbation

The proposed method was validated using the implemented software system with data from the Top50 ranking as well as international supercomputer lists. Examples of the operation of key analytical components of the interface are presented below, demonstrating the practical applicability of the developed approach.

4.1. Data Collection and Administrative Dashboard

The validation of the data collection workflow demonstrates the end-to-end sequence of operations: starting with the entry of a system's inclusion via the administrative panel, followed by the verification of performance metrics against source documentation prior to final recording in the database. The interface streamlines this process by enabling the simultaneous selection of both the target system and the relevant ranking, while automatically populating performance unit fields in accordance with the measurement standards used in the selected list. Furthermore, the workflow supports the extensibility of core reference directories such as rankings and measurement units within a unified, coherent administrative environment.

To facilitate the addition of extensions, a dedicated management panel is provided, conveniently accessible directly beneath the ranking selection list. The introduction of new ranking lists also entails the definition of associated benchmarks used for system evaluation and ordering, including the option to register entirely new benchmark types as the field evolves. Similarly, a separate but logically linked panel is provided for the expansion of performance measurement units, ensuring that the system can adapt to emerging metrics without disrupting the core data model or requiring structural schema changes.

4.2. Approbation of the Lifelines of a Separate System

On the page of a specific system, the visualization of its lifeline has been validated, a graphical representation of the sequence of its entries in international rankings. The interface supports filtering trajectories by individual rankings, interactive tooltips displaying entry details (position, release date, performance metrics), and a tabular summary of all recorded observations for verification and export in CSV format. Figure 3 shows a portion of the table listing the Lomonosov-2 system's entries in the TOP500 ranking.

As an example, the Lomonosov-2 system is considered [20], which has an extensive history in the Top500 and Green500 rankings and is also referenced in other lists.

Rating	Position	Edition	Edition Publication Month	Performance Metrics
Top 500	130	44	2014-11-01	Rmax: 319.8 TFLOP/S, priority: 1, Rpeak: 423.42 TFLOP/S, priority: 2
Top 500	23	45	2015-06-01	Rmax: 1849.0 TFLOP/S, priority: 1, Rpeak: 2575.87 TFLOP/S, priority: 2

Figure 3. Lomonosov-2 System Entries in the Top500 Ranking (table excerpt)

Each ranking list is rendered as a distinct, self-contained table, ensuring modular data presentation and facilitating independent analysis of different datasets. Figure 4 shows the second table summarizing the Lomonosov-2 systems appearances in the GREEN500 ranking. The system implements a role-based access control mechanism, wherein authorized administrators are granted privileged operations over the data entries. Specifically, administrators can edit existing records to correct metadata or update performance metrics, as well as remove obsolete or erroneous entries when necessary. This administrative layer is designed to maintain data integrity and currency while preserving an audit trail of modifications, thereby supporting the long-term reliability of the ranking repository.

Rating	Position	Edition	Edition Publication Month	Performance Metrics
GREEN 500	128	9	2017-06-01	Energy Efficiency: 1.948 GFlops/watts, priority: 1
GREEN 500	130	10	2017-11-01	Energy Efficiency: 1.948 GFlops/watts, priority: 1

Figure 4. Lomonosov-2 System Entries in the GREEN500 Ranking (table excerpt)

Interactive tables provide a detailed representation of the dataset, offering access to historical records of system inclusions and their corresponding performance metrics across multiple ranking releases. This granular approach is essential because standard HPC rankings often represent merely the “tip of the iceberg”, frequently lacking precise longitudinal data and comparative capabilities. To address the demand for novel data presentation formats and to revitalize traditional ranking lists, a lifecycle diagram is positioned directly above the tables, Fig. 5. This visualization serves as a critical instrument for tracking the evolution of the HPC fleet over time, enabling users to analyze system presence duration in rankings a factor increasingly associated with research on system upgradability and the long-term growth of regional computational capacity.

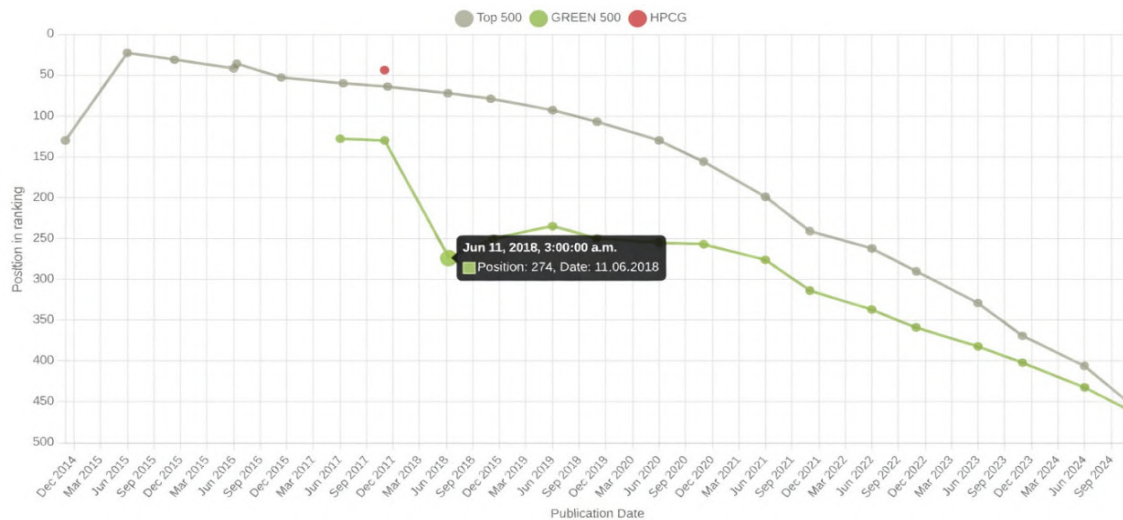


Figure 5. Lomonosov-2 system's lifeline on the system's page in the Top50 rating

4.3. Aggregated Display of Lifelines in the Statistics Section

In the Statistics section, validation of the aggregated representation of lifelines for all Russian systems across the entire observation period has been conducted. In practice, each system is displayed as an individual trajectory, which enabled assessment of position distribution density, identification of periods of strengthening or weakening presence of Russian systems in international rankings, as well as visual comparison of the synchronicity of their movement within the ranking space. The obtained graphical representations confirmed the clarity and informativeness of the proposed approach for analyzing the dynamics of Russian system entries into global rankings.

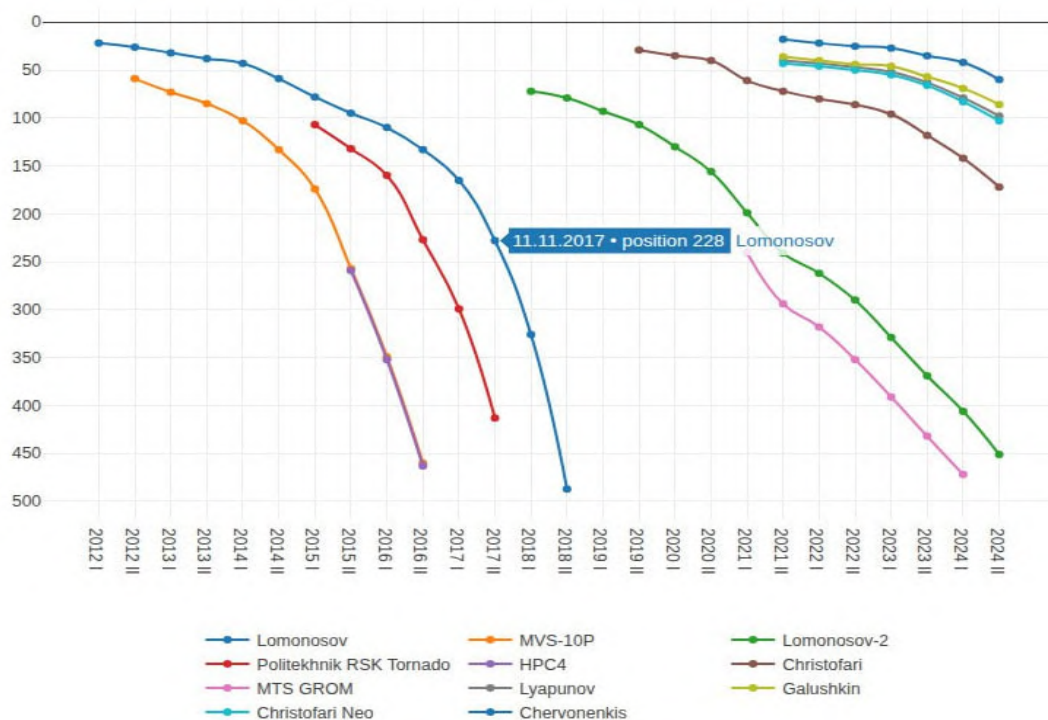


Figure 6. Lifelines of the Top50 participants mentioned in the Top500 in the statistics section

As part of this approbation, the mapping of the lifelines of Russian systems in the Top50 and Top500 ratings in a single visual space was also developed. This view made it possible to simultaneously trace the trajectories of the systems both on the international list and on the domestic rating field, identifying general trends in their advancement, periods of divergence and convergence of positions, as well as features of synchronicity or, conversely, asynchrony of dynamics in the two hierarchies. Figure 6 provides an aggregated view of the lifespans of all domestic systems listed in the TOP500. The practical value of this approach lies in the ability to more clearly assess the relationship between changes in the positions of systems in the global ranking and their status in the national context. The combined display of lifelines contributes to a deeper analysis of transitions between levels of rating significance, allows you to record stable and short-term changes in rank position and thereby increases the informative presentation of the results in the framework of comparative research. Figure 7 provides an aggregated view of the lifespans of all domestic systems participating in the TOP500, as well as a simultaneous display of the trajectories of the systems in the Top50.

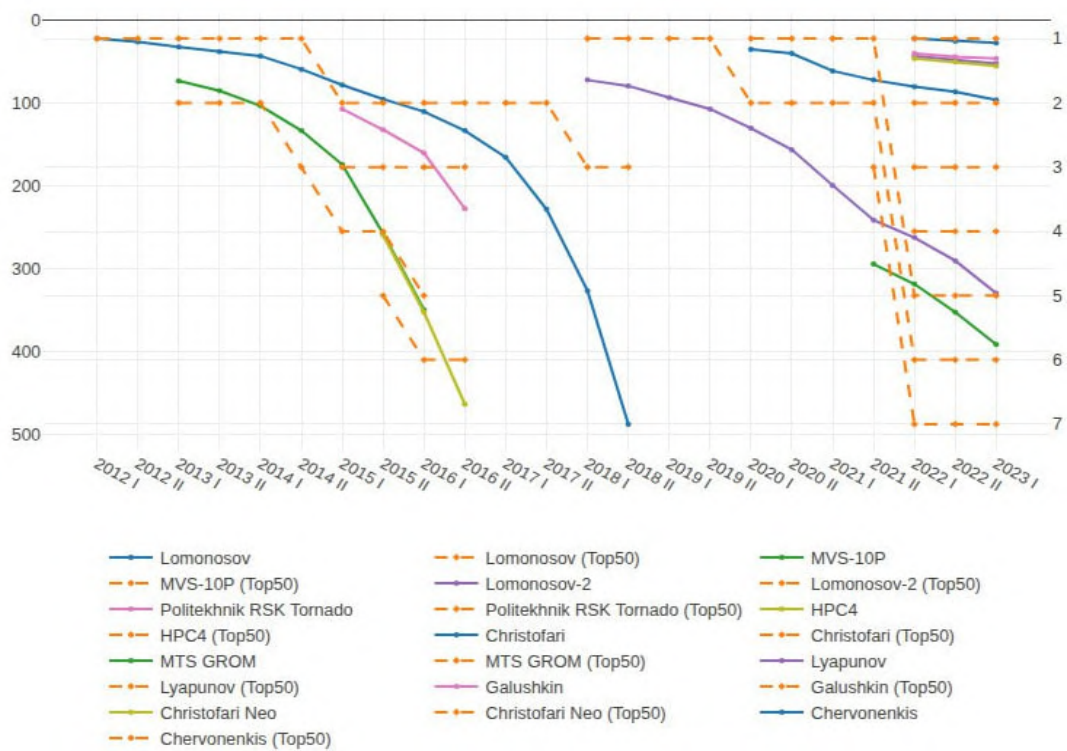


Figure 7. Joint mapping of the trajectories of the presence of Russian systems in the Top500 and Top50 rankings

4.4. Displaying Relative System Performance Indicators

The display of temporal trajectories for normalized indicators has been approved: the ratio of achieved LINPACK performance and the ratio of theoretical peak performance. Figure 8 shows the relative peak and theoretically achievable performance ratios for the Lomonosov-2 system. The graphical representation facilitates cross-release comparison of relative metric values and temporal dynamics analysis. Concurrent visualization enables assessment of the gap between actual and theoretical performance, supporting conclusions regarding computational system implementation efficiency throughout observation periods.



Figure 8. Displaying time series of normalized indicators: the percentage of LINPACK performance achieved and the percentage of the theoretical peak value for the Lomonosov-2 system

4.5. Appropriation of the Display Technique in the “Japanese Candlestick’s” Format

During validation, the visualization of position dynamics in the “candlestick” format was examined not only in terms of technical correctness but also from the perspective of its analytical usefulness. For each ranking release, a candle is constructed reflecting the range of positions of Russian systems, where the open value corresponds to the best observed position and the close value to the worst. This representation enables a compact yet informative depiction of positional variability within each release, allowing for quick identification of periods characterized by either concentration of systems in narrow ranking intervals or, conversely, significant dispersion across a wider range of positions. The visualization is presented as an interactive graph (Fig. 9), where each element corresponds to a separate observation period. In the default view, half-year intervals are displayed, which corresponds to the frequency of publication of the TOP500 ranking.

The correctness of boundary value calculations for the candles and the correspondence of displayed data to database records were thoroughly confirmed through comparison with the underlying dataset. In addition, particular attention was paid to the stability and consistency of the visualization when applied to different temporal segments, ensuring that the graphical representation remains accurate regardless of the selected observation window. Interactive presentation features were also verified, including the ability to obtain an initial overview across the entire observation period, switch to an arbitrary time interval directly within the graph area, and display detailed per-release information via tooltips. Taken together, these validation results demonstrate that the proposed visual tool is not only technically reliable but also effective for both rapid exploratory analysis and detailed verification of ranking position values.

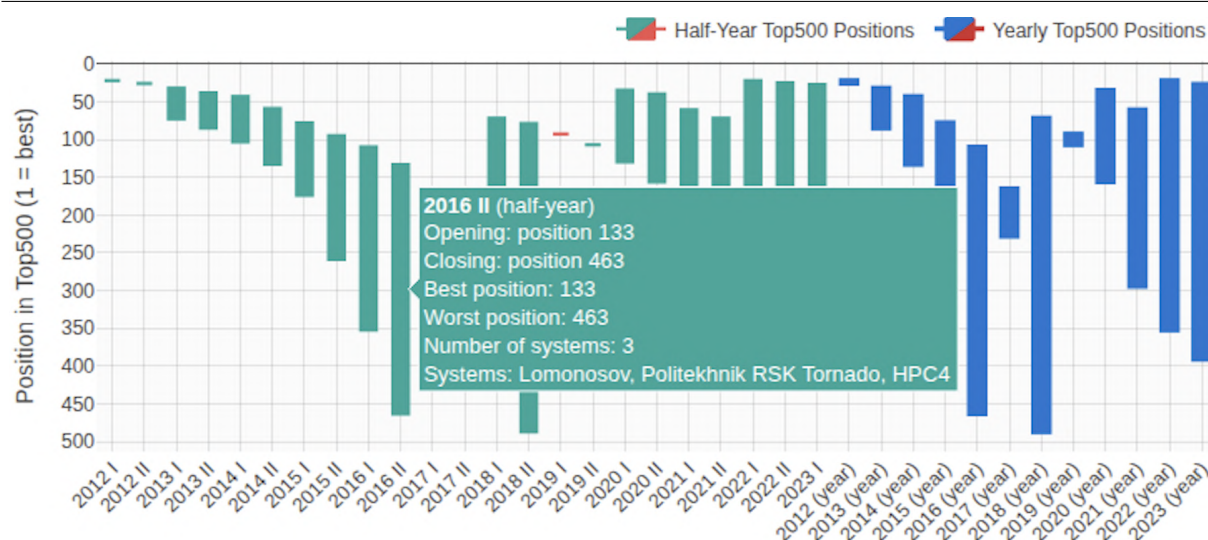


Figure 9. Approbation of the Japanese Candle Technique

4.6. Review of Transition Matrices between Rating Sectors

During validation, the calculation of system transition matrices between sectors of the ranking space (1–10, 11–20, etc.) was performed on the basis of actual observations collected from the integrated dataset. The empirical frequencies of transitions between categories were first computed and then transformed into relative transition probabilities in order to analyze the structural dynamics of the systems positions in the ranking space.

Special attention was paid to verifying the correctness and stability of the calculated probabilities. In particular, the results obtained for selected time intervals were examined to ensure that the normalization of transition frequencies was performed correctly and that the resulting probability distributions were consistent with the observed transitions. Additional verification procedures included cross-checking the aggregated transition counts with the underlying dataset and confirming that the computed probabilities accurately reflected the empirical structure of movements between ranking sectors.

Taken together, these capabilities enable a more comprehensive interpretation of the dynamics of supercomputer systems within the ranking space. The combination of empirical transition counts and normalized probabilities makes it possible to analyze not only the magnitude of movements but also the structural patterns underlying the redistribution of systems across different ranking sectors over time. Figure 10 displays the transition matrices in the users default view. Figure 11 illustrates the users ability to enter a dynamic date range to select the study interval for which the probabilistic transition matrices will be constructed.

The proposal to display numerical values of transitions from one group of the rating space to another is particularly important, since it allows us to move from a qualitative description of the dynamics to its more accurate quantitative interpretation. In contrast to simply recording the fact that systems move between rank intervals, numerical transition indicators make it possible to assess the intensity of such changes, identify dominant migration trends, and determine which transitions are stable and which are episodic. This provides a deeper understanding of the structural dynamics of the distribution of systems by rating sectors.

In addition, the representation of transitions in numerical form increases the visibility and comparability of results between different time periods. This is especially useful when analyzing

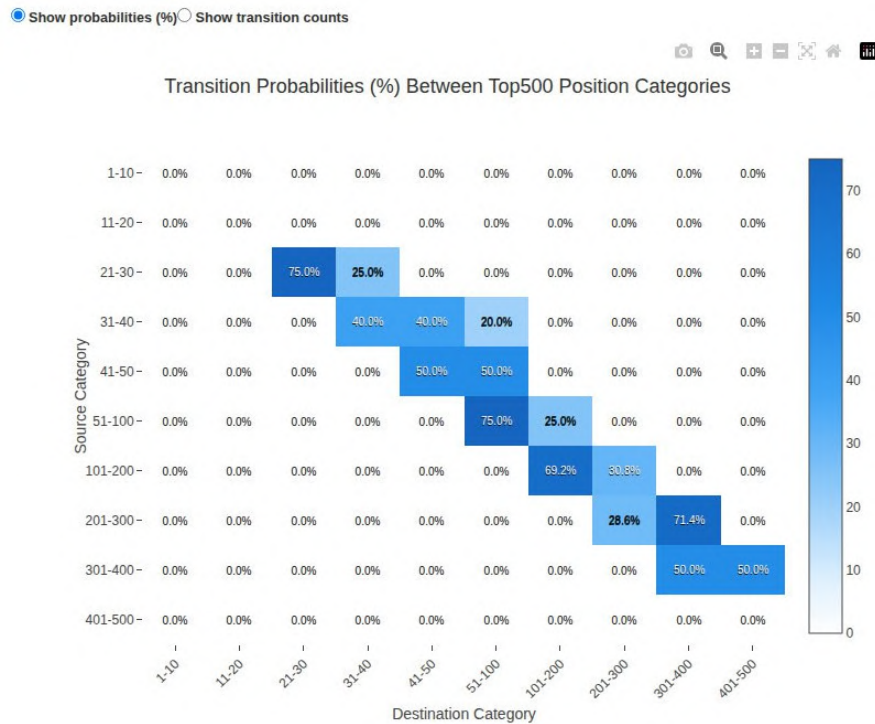


Figure 10. Matrix for analyzing the transition of supercomputers between sections of the list over a selected interval



Figure 11. Matrix for analyzing the transition of supercomputers between sections of the list over an arbitrary time interval

the evolution of the rating position of systems, as it allows you to track not only the fact of displacement, but also the scale of redistribution between groups. As a result, it becomes possible to identify periods of active restructuring of the composition of rating intervals, as well as to assess the degree of stability of the position of individual systems in dynamics. Figure 12 shows how users can view the number of system transitions between rating categories, as well as how they can select a date range spanning two years.

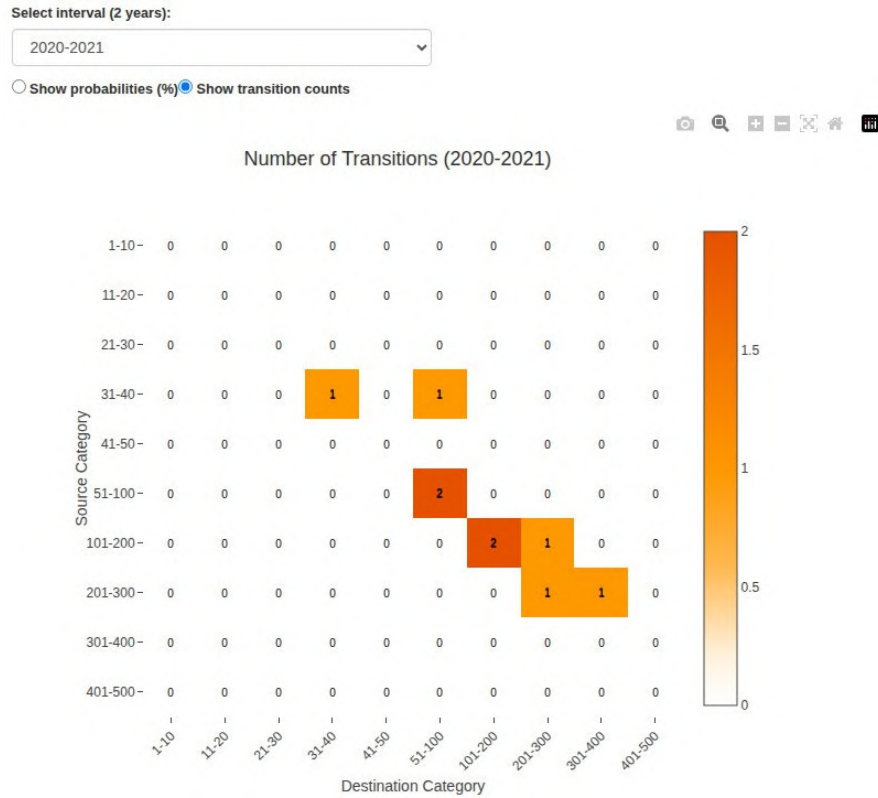


Figure 12. Displaying the number of transitions

Conclusions and Practical Significance

This article presents a new methodological approach and a set of software tools for an in-depth analysis of the dynamics of Russian supercomputer systems presence in global rankings. Users can explore the functionality on the Top50 ranking website, where it is available. Unlike traditional methods based on snapshots of publications, the proposed “lifelines” concept shifts the focus of the research to analyzing the life cycles of specific physical complexes based on their development trajectories. The principal results of the study are as follows:

- An extensible data model integrated into the *Top50* platform was developed, which ensures correct versioning of systems and enables comparison of their characteristics across different editions of international lists (e.g., *Top500*, *Graph500*, *Green500*, etc.).
- The rationale for using normalized indicators relative to the national ranking is substantiated. This normalization made it possible to reveal latent technological trends and to assess the actual competitiveness of domestic systems that are often statistically insignificant at the global scale.
- A suite of advanced visualization and analysis tools was implemented, including transition matrices and candlestick charts. Method validation demonstrated that these tools effec-

tively diagnose the stability of system positions and support forecasting of technological obsolescence scenarios.

The practical significance of the work is confirmed by the deployment of a functional analytical extension to the Top50 website, which serves as a convenient instrument for experts and decision-makers when evaluating the effectiveness of investments in national HPC infrastructure.

Upon analyzing the obtained images, distinct trends were identified and patterns were discovered:

- The lifelines of systems installed between 2020 and 2021 have a lower slope than those of systems installed between 2014 and 2016. This indicates a slower rate of degradation compared to previous-generation systems.
- The transition matrices show that the ranges 31–40 and 41–50 are zones of increased instability. Systems in these ranges rarely retain their original category in subsequent releases. Most of them begin a gradual shift toward lower ranges. These transitions occur smoothly over several releases rather than abruptly, indicating gradual technological obsolescence rather than a sudden loss of relevance. The observed transition probabilities in this range suggest that the 31–50 range is the area where decisions about the system’s future are made, since maintaining its position will require significant modernization.
- The 101–200 ranking segment demonstrates the most stable pattern of transitions and can be viewed as a zone of relative stability for computing systems, where maintaining positions can be achieved without significant modernization, striking a balance between sufficient computing power and the cost of the upgrade.
- “Japanese candlesticks” and the percentage of updated systems metric reveal a three-tiered structure of the Top500 ranking: Positions 1–50 (high turnover/race), 101–200 (stagnation), 201–500 (rapid turnover/budget cycles), where the ownership strategy is dictated by the economics of system operation.

Acknowledgements

The results were obtained in Lomonosov Moscow State University with the financial support of the Russian Science Foundation (agreement No. 25-11-00181).

This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.

References

1. Benhari, A., Denneulin, Y., Desprez, F., Dufossé, F., Trystram, D.: Green HPC: An analysis of the domain based on Top500. <https://hal.science/hal-04519645> (2024), accessed: 2026-04-05
2. Elishakov, D.G., Nikitenko, D.A.: Updateability of systems and their components in the Top50 supercomputer ranking. In: Parallel Computational Technologies – XIX All-Russian Conference with International Participation, PaVT’2025, Moscow, April 8–10, 2025. Short Papers and Poster Descriptions. pp. 335–335. South Ural State University Publishing, Chelyabinsk (2025). <https://doi.org/10.14529/pct2025>

3. Feitelson, D.G.: On the interpretation of Top500 data. *Int. J. High Perform. Comput. Appl.* 13(2), 146–153 (1999). <https://doi.org/10.1177/109434209901300204>
4. Khan, A., Sim, H., Vazhkudai, S.S., Butt, A.R., Kim, Y.: An analysis of system balance and architectural trends based on Top500 supercomputers. In: *Proceedings of the International Conference on High Performance Computing in Asia-Pacific Region (HPCAsia 2021)*. pp. 11–22. ACM (2021). <https://doi.org/10.1145/3432261.3432263>
5. Nikitenko, D., Antonov, A., Zheltkov, A., Voevodin, V.: Describing HPC system architecture for understanding its capabilities. In: Voevodin, V., Sobolev, S. (eds.) *Supercomputing. RuSCDays 2020. Communications in Computer and Information Science*, vol. 1331, pp. 425–435. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-64616-5_37
6. Nikitenko, D., Zheltkov, A.: The Top50 list vivification in the evolution of HPC rankings. In: Sokolinsky, L., Zymbler, M. (eds.) *Parallel Computational Technologies. Communications in Computer and Information Science*, vol. 753, pp. 14–26. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67035-5_2
7. Nikitenko, D., Zhumatiy, S., Paokin, A., et al.: Evolution of the Octoshell HPC center management system. In: Sokolinsky, L., Zymbler, M. (eds.) *Parallel Computational Technologies. Communications in Computer and Information Science*, vol. 1063, pp. 19–33. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-28163-2_2
8. Nikitenko, D.A.: Top 50: Ranking of the most powerful computing systems in the CIS. In: Voevodin, V.V., Tyrtshnikov, E.E. (eds.) *Numerical Methods, Parallel Computing and Information Technologies: Collection of Scientific Works*, pp. 173–182. Moscow State University Publishing, Moscow (2008)
9. Nikitenko, D.A.: Hierarchical model of architecture of supercomputer systems for comparison and ranking. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering* 11(4), 5–18 (2022). <https://doi.org/10.14529/cmse220401>
10. Steed, C.A.: Supercomputing Top500 visualizations. <https://csteed.com/Top500Vis/> (2026), accessed: 2026-04-05
11. Strohmaier, E., Meuer, H.W., Dongarra, J., et al.: The Top500 list and progress in high-performance computing. *Computer* 48(11), 42–49 (2015). <https://doi.org/10.1109/MC.2015.338>
12. The GRAPH500 Organization: The GRAPH500 List. Official website. <https://graph500.org/> (2026), accessed: 2026-04-05
13. The GREEN500 Organization: The GREEN500 List. Official website. <https://green500.org/> (2026), accessed: 2026-04-05
14. The IO500 Organization: The IO500 List. Official website. <https://io500.org/> (2026), accessed: 2026-04-05
15. The Linux Foundation: 20 years of Top500 data show Linux’s role in supercomputing breakthroughs. <https://www.linuxfoundation.org/blog/blog/>

- 20-years-of-top500-data-show-linuxs-role-in-supercomputing-breakthroughs (2013), accessed: 2026-04-05
16. The PostgreSQL Global Development Group: PostgreSQL documentation. <https://www.postgresql.org/docs/> (2026), accessed: 2026-04-05
 17. The Ruby on Rails Team: Ruby on Rails guides. <https://guides.rubyonrails.org/> (2026), accessed: 2026-04-05
 18. The Top500 Organization: The Top500 List. Official website. <https://top500.org/lists/top500/> (2026), accessed: 2026-04-05
 19. Top50 Project: The Top50 Supercomputers List. Official website. <https://top50.parallel.ru/list> (2026), accessed: 2026-04-05
 20. Voevodin, V.V., Antonov, A.S., Nikitenko, D.A., et al.: Supercomputer Lomonosov-2: Large scale, deep monitoring and fine analytics for the user community. *Supercomputing Frontiers and Innovations* 6(2), 4–11 (2019). <https://doi.org/10.14529/jsfi190201>
 21. Voevodin, V.V., Chulkevich, R.A., Kostenetskiy, P.S., et al.: Administration, monitoring and analysis of supercomputers in Russia: a survey of 10 HPC centers. *Supercomputing Frontiers and Innovations* 8(3), 82–103 (2021), <https://doi.org/10.14529/jsfi210305>