International Journal of Asian Business and Information Management
Volume 15 • Issue 1

A Study on the Operational Efficiency of National Supercomputer Resources Based on DEA Model:
Focusing on Specialized Center Resources by Field

Hyungwook Shim, Korea Institute of Science and Technology Information, South Korea*
https://orcid.org/0000-0001-8688-0309

Jaegyoon Hahm, Korea Institute of Science and Technology Information, South Korea

ABSTRACT

As a measure to improve the operational efficiency of the Korean supercomputer joint utilization resources, an efficiency analysis using DEA was conducted to derive the improvement level of the input and output factors of the operation plans of seven specialized centers. The factors that constitute the objective function of efficiency were identified through the operation plans submitted by the specialized centers, and DEA analyzed the improvement factors and levels of inefficient specialized centers through the CCR and BCC models, and then conducted an additional super-efficiency analysis to compare the relative efficiency levels among the efficiency groups. As a result of the analysis, the inefficiency of specialized centers in the meteorological/climate/environmental fields was identified from the reference group, and the need for improvement and the level of improvement of input and output variables were identified. In addition, the authors suggested ways to improve the establishment of resources and operational targets, such as a joint utilization ratio to secure financial efficiency at the government level in establishing a joint utilization resource.

KEYWORDS

BCC, CCR, DEA, operational efficiency, supercomputer

INTRODUCTION

Supercomputing resources are a key infrastructure that determine a country's technological competitiveness, and it is crucial to have adequate resources. In the Top500, a biannual ranking of supercomputers, the top ten positions change every year due to the race to build large-scale supercomputers, and exascale supercomputers have emerged in just a few years since the advent of petascale supercomputers.

DOI: 10.4018/IJABIM.336561
*Corresponding Author

However, in recent years, many countries that do not have the original technology to manufacture supercomputers have adopted a system of sharing multiple small- and medium-sized infrastructures as a single resource instead of building new infrastructures with huge budgets. The sharing economy paradigm has been applied to supercomputer operations. The Republic of Korea has also started to establish a joint utilization system by designating seven specialized centers in 2022 and sharing computing resources among them according to user demand.

Each year, each center determines the proportion of resources for joint utilization from its total resources and utilizes them as computational resources in other fields. The existing resources and resources from other fields are added together to make up for the lack of resources.

Each year, each center submits the ratio to the government for approval. To date, the ratio has been set at an appropriate level by the specialized centers based on their own demand surveys, but the need for more effective joint utilization is emerging due to the continuous increase in supercomputing demand.

For these reasons, this paper applies the method of using the relative efficiency of the operation planning factors of the specialized centers as an optimization method for the most efficient operation of the joint utilization resources. It identified common input and output variables among specialized centers and conducted efficiency analysis using CCR, BBC, and super-efficiency mode (SEM) models. Using the analysis results, we derived the input and output variables that can be improved and the level of improvement to improve the efficiency of inefficient centers.

This paper consists of six chapters. Chapters 1 and 2 present academic value through a qualitative analysis of the background, research purpose, and major theories. Chapter 3 introduces the joint utilization system and describes the operation plan of each specialized center. Chapter 4 presents the research model and describes the research procedures and methods. Chapter 5 presents the results of the efficiency analysis, explains the method and type of research data, and presents descriptive statistics and correlation analysis results. And then, the efficiency analysis results of the CCR, BCC, and SEM models are presented. Finally, in Chapter 6, the results are summarized and the viewpoint was straightened out, and the final point and pursuit plan of this paper is presented.

RESEARCH AND RELATED THEORIES

Prior Research Analysis

In Thore (1996), data envelopment analysis (DEA) was applied to rank the efficiency of United States (US) computer companies during a 10-year period. To reflect the dynamic setting of the computer industry, the inputs included investment in real capital and expenditures on R&D; the outputs were sales revenues, profits, and market capitalization. It developed a procedure for studying the time path of the observed DEA ratings of a high-tech company over its product cycles (Thore, 1996).

Storto (2015) presented a method useful for measuring a supercomputer's performance. This method was based on the implementation of DEA. It used different formulations of the cross-efficiency concept to calculate a comprehensive efficiency index to rank supercomputers. The method was adopted to perform a benchmarking study of the sample, including 77 supercomputers utilized to solve complex problems in research applications (Storto, 2015).

Wang et al. (2021) aimed to assess the performance efficiency of cloud computing service providers in the United States of America by applying the data envelopment analysis models. The efficiency of cloud computing providers was evaluated based on the assumption of the noncooperative game among cloud computing providers in which providers selfishly choose the best strategy to maximize their payoff with three stages (Wang et al., 2021).

Raja and Ramaiah (2016) built up a consumer and cloud-data envelopment analysis (CCDEA) trust assessment model for evaluating cloud services in two stages. In the first stage, the believability index of each cloud consumer (C) was calculated. The second stage incorporated a cloud-data envelopment analysis (C-DEA) model for the trust assessment of cloud services from the viewpoint of cloud consumers (Raja & Ramaiah, 2016). Kao et al. (2017) intended to develop alternative network data envelopment analysis (NDEA) models for evaluating cloud service businesses.

By considering various internal functions and processes of the services in multiperiod settings, we designed three evaluation models: (1) dynamic black-box data envelopment analysis (DBDEA); (2) static network data envelopment analysis (SNDEA); and (3) dynamic network data envelopment analysis (DNDEA). Using multiobjective programming (MOP) techniques, the three NDEA models are formulated and solved for the cloud service industry.

An empirical study was conducted to evaluate the performance of the cloud service industry (Kao et al., 2017). Leonenkov (2019) used the historical data from the two largest Russian supercomputers to create a number of metrics in order to provide the definition of resource management "efficiency." The data from both Lomonosov and Lomonosov-2 supercomputers consisted of over 1 year of history of job executions. Lomonosov and Lomonosov-2 efficiency in terms of CPU hours utilization was considerably high, nevertheless, our global goal is to offer a way to maintain or improve this metric when maximizing others examined in the paper (Leonenkov & Zhumatiy, 2019).

Mamaeva et al.'s (2018) paper presented three approaches to study the efficiency of supercomputer resource usage based on monitoring data analysis. The first approach performed an analysis of computing resource utilization statistics, which allows to identify different typical classes of programs, to explore the structure of the supercomputer job flow, and to track overall trends in the supercomputer behavior. The second approach was aimed specifically at analyzing off-the-shelf software packages and libraries installed on the supercomputer since the efficiency of their usage is becoming an increasingly important factor for the efficient functioning of the entire supercomputer. Within the third approach, abnormal jobs—jobs with abnormally inefficient behavior that differ significantly from the standard behavior of the overall supercomputer job flow—are being detected (Mamaeva et al., 2018).

Although there are not many analogous studies in the supercomputer field, efficiency studies using DEA have been conducted continuously. These have been categorized as investment efficiency, performance efficiency, operational and service efficiency. However, there has not been much efficiency research on the new trend of joint utilization supercomputer resources. There are several reasons for this, including the need to develop the technology first and the fact that the concept of joint utilization is biased toward the service concept. In the Republic of Korea, on the other hand, joint utilization is a government-led management system in which the government designates participating organizations and approves operational goals. Therefore, in order for the system to be recognized as sustainable, it is most important to secure the efficiency of government financial support, and for this purpose, efforts are needed to improve the efficiency of the system unit as well as the efficiency of individual participating organizations.

Research Need and Practical Contributions

This is the beginning of a pioneering scheme for jointly utilizing supercomputing resources. Efficient operation and effective performance creation of the specialized centers that make up this system are very important, and an initial operation plan that focuses on setting achievable goals and resource investment plans must be prepared. However, because this system has very few operational cases internationally, there are few references to refer to for establishing an operation plan, and there is a lack of basis for establishing achievable goals.

As an alternative to solve this problem, this paper conducted an analysis of a DEA's operational efficiency. It is common for existing DEAs to analyze efficiency based on the performance and results derived from the Decision Making Unit (DMU). This will be used to improve the DMU's efficiency in the next phase and year. However, on the contrary, we applied it as a new method of setting input and output variables to achieve stable goals of the DMU based on the operation plan, not the operation results of the DMU.

The practical contributions of this paper are as follows. First, not much research has been done on supercomputer joint utilization systems compared to other fields. Furthermore, research on the operational aspect rather than the technical aspect is even more poor. Therefore, this paper establishes key prior references and can be used as basic data for follow-up research. Additionally, in the future, many experts will be able to use DEA in a variety of ways with a broader perspective, rather than limiting it to just interpreting results.

Theory of Efficiency Analysis

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors*).

DEA models are divided into several models depending on the assumptions and perspectives on efficiency, and the most widely used models are the constant returns to scale (CRS) model, called CCR, and the variant returns to scale (VRS) model, called BBC. These models are divided into multiplier models, ratio models, and envelope models according to the optimization method of the objective function representing efficiency, and this paper uses an envelope model that provides various information in terms of interpreting the results.

First, the input-directed CCR envelope model is analyzed by focusing on slack, which implies additional room for improvement to reach the state above the efficiency boundary. It can be described as a model that finds the smallest input level θ by reducing all m inputs by a certain percentage while maintaining an output level that is at least equal to or greater than the current output level. A mathematical representation of this model in terms of linear programming would look like Equations (1)–(4). x, y are the inputs and outputs of a particular DMU, and λ is the weight assigned to each DMU. After creating an efficiency boundary using the linear combination of inputs $\sum_{j=1}^{n} x_{ji}\lambda_j$ and outputs $\sum_{j=1}^{n} y_{jr}\lambda_j$, we impose the constraint that the output of a DMU is less than or equal to the output of this efficiency boundary. And the input is constrained to be greater than or equal to the input above the efficiency boundary. At this time, the value that reduces the distance from the efficiency boundary is calculated as the efficiency score θ.

$$Min\ \theta \tag{1}$$

$$s.t.\ x_{ki}\theta \geq \sum_{j=1}^{n} x_{ji}\lambda_j \left(i = 1,2,\cdots,m\right) \tag{2}$$

$$\sum_{j=1}^{n} y_{jr}\lambda_j \geq y_{kr} \left(r = 1,2,\cdots,s\right) \tag{3}$$

$$\lambda_j \geq 0. \left(j = 1,2,\ \cdots,n\right). \tag{4}$$

In other words, the objective function of the DMU is the problem of minimizing θ. If $\theta = 1$, the DMU is efficient because there are no more inputs to reduce, and if $\theta < 1$, the DMU is inefficient because there is room to reduce inputs.

Next, we need to maximize the sum of the input and output slack while maintaining the efficiency level. This is done by using the current input and output levels and the difference between input and output above the efficiency boundary to satisfy the following conditional equations (Equations (5) – (7)). The margins of the i-th input and j-th output factors are expressed as s_i^-, s_r^+. If θ is 1 and w is 0, it is judged as a DMU that is efficient and satisfies strong efficiency with no slack.

$$Max\ w = \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \tag{5}$$

$$s_i^- = x_{ki}\theta^* - \sum_{j=1}^{n} x_{ji}\lambda_j \tag{6}$$

$$s_i^+ = \sum_{j=1}^{n} y_{jr}\lambda_j - y_{kr}\ . \tag{7}$$

Second, the input-directed BBC envelope is given by Equations (8)–(12). Compared to the CCR envelope, the constraint $\sum_{j=1}^{n}\lambda_j = 1$ is added.

$$Min\ \theta_k - \epsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right) \tag{8}$$

$$s.t.\ \theta x_{ik} - \sum_{j=1}^{n} x_{ji}\lambda_j - s_i^-\ \left(i = 1, 2, \cdots, m\right) \tag{9}$$

$$y_{rk} - \sum_{j=1}^{n} y_{jr}\lambda_j + s_r^+\ \left(r = 1, 2, \cdots, s\right) \tag{10}$$

$$\sum_{j=1}^{n}\lambda_j = 1 \tag{11}$$

$$\lambda_j \geq 0,\ \ s_i^- \geq 0,\ \ s_i^+ \geq 0\ . \tag{12}$$

Imposing the constraint $0 \leq \sum_{j=1}^{n}\lambda_j \leq 1$ assumes a declining return to scale (DRS) and imposing $\sum_{j=1}^{n}\lambda_j \geq 1$ assumes an increasing return to scale (IRS). This is summarized in Table 1 (Novikov et al., 2016; Niu et al., 2013; Atta Mills et al., 2020).

The scale efficiency (SE) is a useful metric that can be obtained from the BBC model and can be defined as the following Equation (13). SE takes on values greater than 0 and less than 1, with values closer to 1 indicating no efficiency losses due to scale. In the case of harvest invariance to scale, the SE is equal to 1, indicating that the DMU is at the most optimal scale level.

$$SE_k = \frac{CCR\left(\theta_k^*\right)}{BCC\left(\theta_k^*\right)}\ . \tag{13}$$

The SEM solves the problem that the previous models had: when many DMUs have an efficiency value of 1, it is not possible to evaluate the relative value between them. The super-efficiency model can be applied to both CCR and BBC models, but only the SEM of the BBC model will be analyzed.

Table 1. Condition of return to scale

Condition	RTS
CCR(θ*) = BCC(θ*)	CRS
CCR(θ*) ≠ BCC(θ*), CCR(θ*) = DRS (θ*)	DRS
CCR(θ*) ≠ BCC(θ*), CCR(θ*) = IRS (θ*)	IRS

JOINT UTILIZATION RESOURCES OPERATION PLAN

A supercomputer specialized center is defined as an institution that possesses expertise for the professional use of supercomputers, provides specialized services based on resources, manpower, and technology specialized in the field, conducts research and development, and promotes the use of supercomputers. The functions and roles of the specialized centers in Table 2 include the establishment and operation of supercomputing resources by field, service provision, base application research and dissemination of research results, large-capacity data management and operation support, and human resource training (Cooper, 2011).

Institutions designated as specialized centers are required to develop a 5-year mid- to long-term operating plan and an annual operating plan. Seven specialized centers have developed and approved their operating plans and are currently in the second year of implementation. The operational plans for each center are as follows.

The space specialized center has established a vision to revitalize basic research related to future space exploration and development by creating a supercomputing infrastructure and strengthening support services. As a major initiative, it plans to build a supercomputing infrastructure for space, and optimization parallelization research for space applications, and provide incubation programs for related organizations to expand the joint utilization resources.

The disaster specialized center plans to build new supercomputer resources to support marine forecasting and disaster response capabilities through supercomputing, as well as the establishment of long-term policies for all coastal and marine areas in the Republic of Korea and to link forecasting data with high accuracy, ultra-high-resolution coastal numerical models.

The specialized center for life/health has set goals to secure supercomputer infrastructure for analysis needs, provide a stable utilization environment, and strengthen technical capabilities. To this end, it plans to build a GPU-based computing environment, establish a joint user technical support system, and promote R&D projects.

The specialized center for meteorology/climate/environment (M/C/E) aims to expand national supercomputing technology based on the Korea Meteorological Administration's supercomputing infrastructure, operation, and utilization technology in the fields and promoting the advancement of supercomputing services by securing infrastructure, providing universal numerical forecasting models and AI-based big data analysis platforms, and supporting the establishment of specialized/unit centers through full-cycle technical support and expanding cooperation.

The specialized center for material/nano plans to establish a joint utilization system for materials/nano supercomputers, support R&D, build and spread an education/research/industrial ecosystem, establish a joint utilization foundation linked to national centers and specialized centers, establish a one-stop total support service for SMEs, and secure core source technologies.

The specialized center for nuclear fusion/accelerators aims to develop simulation, virtualization, and data utilization technologies for nuclear fusion energy development and accelerator device research. The center plans to develop large-scale parallel simulation technology for simulating nuclear fusion phenomena, predict and analyze KSTAR and ITER experiments, support KSTAR and ITER

Table 2. Functions and roles of the specialized centers

	Main content
Function and role	**• Providing supercomputing services in specialized fields • Research and development of supercomputing technology and application specialized in specialized fields • Expansion of resource sharing through participation in the joint utilization system**
Service	• Establishment, operation, and service of specialized resources for each sector centered on ten strategic fields • Utilization support and consulting through field experts • Community development and education by field

device operations and experiments through nuclear fusion virtualization technology, and support simulation technology for accelerator device researchers.

RESEARCH MODEL

Research Procedure and Method

The research procedure is shown in Figure 1. Based on the data from the Ministry of Science and ICT, we surveyed the specialized centers participating in the joint utilization system and analyzed the operation plan data for each specialized center. Through the operation plan, operational factors common to all specialized centers are derived, and input and output variables are selected from the results. The efficiency analysis starts with a review of the appropriateness of factor selection through correlation analysis. Then, CCR and BBC analyses are performed for input and output directions to check efficiency values and distinguish between DRS and IRS based on the assumption of scale profit variability.

Next, the BCC efficiency value is divided by the CCR efficiency value to check whether scale efficiency is achieved. Next, a factor contribution analysis is conducted. Through this, we observe the degree to which input and output variables contribute to the efficiency value, and through the improvement potential analysis, we calculate the value that each variable can be improved to reach the relative target value in the inefficient DMU. Finally, to overcome the limitation that it is difficult to determine relative efficiency, the efficiency value is recalculated through the super-efficiency analysis to compare efficiency among efficient DMUs.

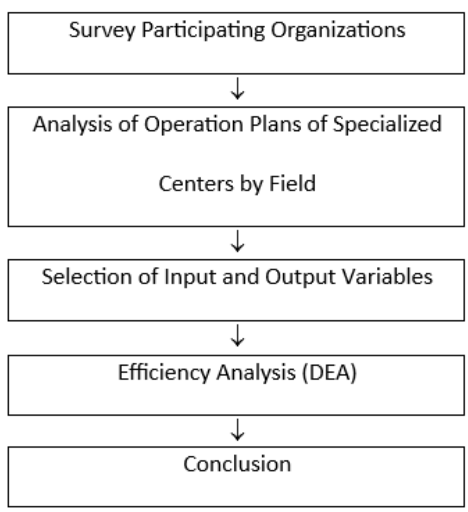
DEA analysis was performed using Banxia Software's Frontier Analyst program, and descriptive statistics and correlation analysis were performed using the IBM SPSS program.

EFFICIENCY ANALYSIS RESULTS

Research Data

To analyze efficiency, the operation plans for each specialized center submitted to the government by the seven specialized centers at the time of designation were referred to. First, the DMU is set to

Figure 1. Research model



seven fields, including space and fusion accelerator, where the centers have been designated so far. Next, variables for input and output were derived by referring to the input resources and performance goals presented by each specialized center. The results for this are shown in Table 3.

In order to set common input and output variables for the seven specialized centers, variables that they had in common, had high importance, and were more suitable for the research purpose were selected. In addition to this, the output variable was further reviewed to see whether it had a connection and causal relationship with the input variable.

The results are shown in Table 4. The input variables are the manpower and budget required to operate the joint utilization resources, and the output variables are the size of supercomputer resources and the proportion of resources allocated for joint use among the resources held by the center.

The research data for the variables are shown in Table 5.

Descriptive Statistics and Correlation Analysis

Descriptive statistics were used to analyze the characteristics of individual groups of variables. Then, correlation analysis was used to quantitatively derive the linear relationship between variables to review and adjust the appropriateness of variable selection. The results of descriptive statistics and correlation analysis are shown in Table 6.

The results of the correlation analysis for DMUs are shown in Table 7. All DMUs were moderately correlated with each other (0.2 to 0.6), but there were no strong correlations above 0.7. In general, the relationship between input and output variables should be positive and statistically significant (Shim et al., 2023). In the case of variable A, we found that this was not the case, but the results were not

Table 3. Total list of variables

DMU	Input	Output
Space	Storage capacity, Project, Operating manpower, Budget, Services, User training program	Resources (new build), Joint utilization resources, Joint utilization rate, Supercomputer utilization rate
Nuclear fusion/ accelerator	Storage capacity, Operating manpower, Budget	Resources (new build), Research performance (paper, etc), Joint utilization resources, Joint utilization rate
Disaster	Operating manpower, Budget	Resources (new build), Joint utilization resources, Joint utilization rate, Institution
Life/health	Operating manpower, Budget	Resources (new build), User satisfaction, research results (papers, patents, etc.), Joint utilization resources, Joint utilization rate, Service safe
Meteorology/ climate/ environment	Operating manpower, Budget, User training program	Resources (new build), User satisfaction, research results (papers, patents, etc.), Joint utilization resources, Joint utilization rate
Material/nano	Operating manpower, Budget, Performance promotion, User training program	Resources (new build), Research performance (paper, etc), Joint utilization resources, Joint utilization rate
Autonomous-driving	Equipment (large simulator, etc.), Operating manpower, Budget, Service	Resources (new build), Joint utilization resources, Joint utilization rate

Table 4. Selected list of variables

	Code	Name	Unit	Note
Input variables	**A**	**Operations staffing**	**Person**	
	B	**Budget**	**$1,000**	
Output variables	C	Resource size	PF	
	D	Joint utilization ratio	%	Joint utilization resources/Total resources

Table 5. Research data

	A	B	C	D
Space	19	64	60	30
Nuclear fusion/accelerator	22	12	30	12
Disaster	10	1,400	70	25
Life/health	17	447	60	31
Meteorology/climate/environment	38	459	32	10
Material/nano	44	66	61	5
Autonomous-driving	12	200	43	14

Table 6. Descriptive statistics results

	N	Min	Max	Mean	Std	Var
A	7	10.00	44.00	23.14	12.97	168.14
B	7	12.00	1,400.00	378.29	486.04	236,230.91
C	7	5.00	31.00	18.14	10.38	107.81
D	7	30.00	70.00	50.86	15.75	248.14

statistically significant. As a result, the variable settings were judged to be reasonable at a certain level, and we proceeded with the analysis without adding or removing variables.

- $p < .05$

Efficiency Analysis

Efficiency Results

The results of the input-oriented efficiency analysis are shown in Table 8. The CCR efficiency was efficient in all but two areas, M/C/E and Material/nano. Of the two inefficient areas, Material/nano was 0.7, which was relatively higher than W/C/E, which was 0.25. For BCC efficiency, only W/C/E was inefficient (0.36), and its reference group was Autonomous-driving and Space. Additionally, the DMU with the largest number of reference DMUs is typically considered the benchmarking DMU, with Autonomous-driving, Space having the second largest number. As for the RTS, W/C/E was IDS, and Material/nano was DRS, while the other sectors had a value of 1, indicating that they were in CRS. Therefore, W/C/E means that the increase in output was greater when the input was increased, so efficiency can be improved by expanding the scale, while Material/nano can be improved by reducing the scale.

Table 7. Correlation analysis results

		A	B	C	D
A	**Pearson correlation**	**1**	**–.409**	**–.693***	**–.247**
B	Pearson correlation	–.409	1	.350	.472
C	Pearson correlation	–.693*	.350	1	.548
D	Pearson correlation	–.247	.472	.548	1

Table 8. Input-Oriented efficiency

DMU	CCR	BBC	SE	RTS	Peers (BBC)
Space	1	1	1	CRS	Space
Nuclear fusion/accelerator	1	1	1	CRS	Nuclear fusion/accelerator
Disaster	1	1	1	CRS	Disaster
Life/health	1	1	1	CRS	Life/health
Meteorology/climate/environment	0.25	0.36	0.69	IDS	Autonomous-driving, Space
Material/nano	0.70	1	0.7	DRS	Materials/nano
Autonomous-driving	1	1	1	CRS	Autonomous-driving

The results of the output-oriented efficiency analysis are shown in Table 9. CCR efficiency was efficient in all but two fields, M/C/E and Material/nano. For BCC efficiency, only M/C/E was inefficient (0.51), and its reference group was Disaster, Materials/nano, and Space. The benchmarking DMUs for the number of references were Disaster, Materials/nano, and Space. For RTS, we can see that M/C/E and Materials/nano were in IRS, while the other disciplines were in CRS. This suggested that M/C/E and Material/nano can improve their efficiency by scaling up.

Input/Output Contributions (BCC)

Input and output contributions quantify the contribution of each input and output variable to the other variables in the efficiency analysis. The contribution of input and output variables for inefficient DMUs is shown in Table 10. In the inefficient DMUs M/C/E, the relative contribution of input variables to the efficiency of input-oriented inputs was 23.4 percentage points higher than that of B, and the relative contribution of output variables was 16.2 percentage points higher than that of D. In contrast, the contribution of output-oriented inputs was higher than that of B. On the other hand, the output-oriented input variable B was 100%, and the output variable was the same, but D was higher, with a very high value of 98.8%. Therefore, the extreme influence of a single variable on the output-oriented efficiency excludes the application of this result.

Peer Contributions

The peer contribution results for the inefficient DMUs are shown in Table 11. In the case of A, Autonomous-driving DMUs had a high contribution for all variables A–D, and characteristically, B was extremely high at 90.4%.

Table 9. Output-Oriented efficiency

DMU	CCR	BCC	SE	RTS	Peers (BBC)
Space	1	1	1	CRS	Space
Nuclear fusion/accelerator	1	1	1	CRS	Nuclear fusion/accelerator
Disaster	1	1	1	CRS	Disaster
Life/health	1	1	1	CRS	Life/health
Meteorology/climate/environment	0.25	0.51	0.49	IRS	Disaster, Materials/nano, Space
Materials/nano	0.70	1	0.7	IRS	Materials/nano
Autonomous-driving	1	1	1	CRS	Autonomous-driving

Table 10. Input/Output contributions (BCC)

DMU	Input oriented		Output oriented	
	Variable	Contributions (%)	Variable	Contributions (%)
Meteorology/climate/environment	Input (A)	61.7	Input (A)	0.0
	Input (B)	38.3	Input (B)	100.0
	Output (C)	41.4	Output (C)	1.2
	Output (D)	58.6	Output (D)	98.8

Table 11. Peer contributions (Weather/Climate/Environment)

	Variable	Contribution (%)
Autonomous-driving	A	65.5
	B	90.4
	C	58.4
	D	68.3
Space	A	34.5
	B	9.6
	C	41.6
	D	31.7

Potential Improvement (BBC)

The potential improvement requirements (%) for input and output variables for inefficient DMUs are shown in Table 12. Considering only the analysis results, it is ideal to reduce both inputs A and B to 63.8% in the input-oriented model and increase both outputs C and D to 97.8% in the output-oriented model compared to the reference group.

In practice, however, this is not always possible due to various constraints. First, the reduction in A is generally difficult to apply in specialized centers. In the case of the Republic of Korea, the pool of supercomputing experts is very limited, and the staffing level included in the operational plan is composed of core employees who have been nurtured through a large investment over a long period of time. In addition, it is difficult to reduce manpower based on operational efficiency alone in the current situation of expanding investment to secure specialized manpower.

On the other hand, the reduction of B is something that the government is also considering, and it is necessary to reduce installation costs that are utilized outside of joint utilization. For example, in many specialized centers, when the joint utilization rate is set at 20%, it is necessary not only to expand the new joint utilization resources, but also to further expand the existing resources. In other words, if an institution with 20 PFs uses 10 PFs (20% of 50 PFs) for joint utilization according to demand, the existing 20 PF resources must be expanded to 50 PFs.

Therefore, it is necessary to set targets based on the number of resources (PFs) alone, rather than establishing operational planning targets based on the existing joint utilization ratio (%), so that the budget can be allocated only to building resources. Next, the output variable C is a level that can be expanded compared to the existing amount of resources, and the output variable D is very high, with an increase of about 63.3% from the existing 32.0%. Therefore, in order to determine the appropriate level, it is necessary to consider the amount of resources (1.9 PF) provided for joint utilization of the existing M/C/E specialized center. In addition, it plans to build 60 PF of new infrastructure, of which 63.3% is about 40 PF. This is about 20 times more than the existing joint utilization resource of 1.9 PF. Therefore, D should be set to an appropriate value that meets the demand by referring to the existing joint utilization resource of each specialized center.

Super Efficiency Model

From the efficiency analysis in the previous section, we can conclude that many of the DMUs have efficient operational plans. However, since we are performing a relative comparison of efficiency across groups, it is meaningful to prioritize the efficiency of the more efficient groups. Therefore, as shown in Table 13, a super-efficiency analysis (SEM) was performed on the efficient DMUs (Mostafa, 2009). In the input-directed SEM, the CCR model showed that Nuclear fusion/accelerator was the most efficient, while the BBC model showed that Life/health and Disaster were the most efficient. In the output-oriented SEM, the results of the CCR model were the same as in the input direction, and the BBC model was the most efficient in four areas: Space, Nuclear fusion/accelerator, Disaster, and Autonomous-driving.

Table 12. Peer potential improvement (BBC) (M/C/E)

	Input oriented			Output oriented		
	Actual	Target	Potential improvement (%)	Actual	Target	Potential improvement (%)
A	38.0	13.8	–63.8	38	25.1	–34.0
B	459.0	166.1	–63.8	459.0	459.0	—
C	10.0	18.0	79.9	10.0	19.8	97.8
D	32.0	47.2	47.6	32.0	63.3	97.8

Table 13. Input-Oriented SEM results

	SE (CCR)	Rank	SE (BBC)	Rank
Space	230.3	2	675.2	2
Nuclear fusion/accelerator	266.7	1	533.3	3
Disaster	195.8	3	1,000.0	1
Life/health	105.2	4	1,000.0	1
Materials/nano	—	—	299.4	4
Autonomous-driving	101.5	5	145.6	5

Table 14. Output-Oriented SEM results

	SE (CCR)	Rank	SE (BBC)	Rank
Space	230.3	2	1,000.0	1
Nuclear fusion/accelerator	266.7	1	1,000.0	1
Disaster	195.8	3	1,000.0	1
Life/health	105.2	4	107.3	2
Materials/nano	—	—	101.6	3
Autonomous-driving	101.5	5	1,000.0	1

CONCLUSION

Discussion of Results and Implications

Given the lack of research on supercomputer joint utilization resources and the fact that they are still in the early stages of implementation, various studies are needed to create the most efficient and effective system. In this context, this study is significant in that it is the first study to identify the input and output variables that constitute the efficiency objective function based on the first operating plans submitted by the seven specialized centers and to derive the efficiency of individual specialized centers by using them. In addition, it is thought that it can be applied as a financial efficiency methodology that the government can utilize institutionally through the operational efficiency of the joint utilization system.

Study Limitations and Future Research

The limitations of this study are that, first, operational efficiency was analyzed at a macro level. From a microscopic perspective, the efficiency of the DMU can be diversified not only by the input/output factors presented here but also by other dimensions or lower forms. Therefore, beyond the analysis based on the current operation plan single data, we plan to derive various factors that are physically input and output factors in the actual operation stage.

Next, there is a lack of related references. Currently, it is difficult to find examples of efficiency studies on institutional operations related to the supercomputer field. Research on resource management has been published recently, but it only targets resources. Therefore, this study should be used as a representative reference in the related field, and to this end, we plan to conduct additional verification of results based on actual operation result data. Beyond academic research, the specialized center plans to investigate the characteristics of each field each year, including resource size and specifications, user demand, research areas, and performance. Through this, we plan to regularly evaluate operational efficiency and prepare improvement plans. The government plans to use the results of this study to establish policies and mid- to long-term implementation plans related to the joint utilization system.

AUTHOR NOTE

This research was supported by the Korea Institute of Science and Technology Information(KISTI) (No. K-23-L02-C03).

REFERENCES

Atta Mills, E. F., Baafi, M. A., Liu, F., & Zeng, K. (2020). Dynamic operating efficiency and its determining factors of listed real-estate companies in China: A hierarchical slack-based DEA-OLS approach. *International Journal of Finance & Economics*. https://onlinelibrary.wiley.com/doi/abs/10.1002/ijfe.1965

Cooper, W. W. (2011). *Handbook on data envelopment analysis*. Springer. https://link.springer.com/book/10.1007/978-1-4419-6151-8

Kao, H. Y., Wu, D. J., & Huang, C. H. (2017). Evaluation of cloud service industry with dynamic and network DEA models. *Applied Mathematics and Computation*, *315*, 188–202. doi:10.1016/j.amc.2017.07.059

Leonenkov, S., & Zhumatiy, S. (2019). Supercomputer efficiency: Complex approach inspired by Lomonosov-2 history evaluation. In V. Voevodin & S. Sobolev (Eds.), *Supercomputing. RuSCDays 2018, Communications in computer and information science* (Vol. 965, pp. 631–640). Springer. doi:10.1007/978-3-030-05807-4_54

Mamaeva, A. A., Shaykhislamov, D. I., Voevodin, V. V., & Zhumatiy, S. A. (2018). Comprehensive efficiency analysis of supercomputer resource usage based on system monitoring data. *Journal of Physics: Conference Series*, *973*, 012068. doi:10.1088/1742-6596/973/1/012068

Mostafa, M. (2009). Benchmarking the US specialty retailers and food consumer stores using data envelopment analysis. *International Journal of Retail & Distribution Management*, *37*(8), 661–679. doi:10.1108/09590550910966178

Niu, D., Wang, R., Zhang, X., & Song, W. (2013). Comparative studies on efficiency of China's regional innovation system on the basis of cooperation measurement model. *Journal of Emerging Trends in Economics and Management Science*, *4*, 147–157. https://hdl.handle.net/10520/EJC135867

Novikov, V. A., Shipulina, L. G., & Sapun, O. L. (2016). *Accounting the scale and synergies in the DEA-analysis.* Semantic Scholar. https://www.semanticscholar.org/paper/Accounting-the-Scale-and-Synergies-in-the-Novikov-Shipulina/08ac124adeafb0cbb559bd2b20a097df69924dea

Raja, S., & Ramaiah, S. (2016). CCDEA (Consumer and Cloud – DEA) based trust assessment model for the adoption of cloud services. *Cybernetics and Information Technologies*, *16*(3), 52–69. doi:10.1515/cait-2016-0034

Shim, H. W., Jung, Y. H., & Hahm, J. G. (2023). A study on the designation institution for supercomputer specialized centers in Republic of Korea. *International Journal of Advanced Computer Science and Applications*, *14*(1), 306–312. doi:10.14569/IJACSA.2023.0140132

Storto, C. (2015). A non-parametric measurement of supercomputers performance. *Advanced Science Letters*, *21*(10), 3257–3260. doi:10.1166/asl.2015.6439

Thore, S., Phillips, F., Ruefli, T. W., & Yue, P. (1996). DEA and the management of the product cycle: The U.S. computer industry. *Computers & Operations Research*, *23*(4), 341–356. doi:10.1016/0305-0548(95)00043-7

Wang, C. N., Nguyen, M. N., Nguyen, T. D., Hsu, H. P., & Nguyen, T.-H.-Y. (2021). Effective decision making: Data envelopment analysis for efficiency evaluation in the cloud computing marketplaces. *Axioms*, *10*(4), 309–326. https://pdfs.semanticscholar.org/2f30/f00e3252aaefe895b5faacbd361513172c88.pdf. doi:10.3390/axioms10040309