



OPEN ACCESS

EDITED BY

Yi Zhang,
University of Technology Sydney, Australia

REVIEWED BY

Gil Speyer,
Arizona State University, United States
Shino Iwami,
NEC Corporation, Japan
Thomas Furlani,
University at Buffalo, United States

*CORRESPONDENCE

Preston M. Smith
✉ psmith@purdue.edu

RECEIVED 16 June 2024

ACCEPTED 15 July 2025

PUBLISHED 21 August 2025

CITATION

Smith PM, Gemmill J, Hancock DY,
O'Shea BW, Snapp-Childs W and
Wilgenbusch J (2025) Application of the
cyberinfrastructure production function
model to R1 institutions.
Front. Res. Metr. Anal. 10:1449996.
doi: 10.3389/frma.2025.1449996

COPYRIGHT

© 2025 Smith, Gemmill, Hancock, O'Shea,
Snapp-Childs and Wilgenbusch. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Application of the cyberinfrastructure production function model to R1 institutions

Preston M. Smith^{1*}, Jill Gemmill², David Y. Hancock³,
Brian W. O'Shea^{4,5,6,7}, Winona Snapp-Childs^{3,8} and
James Wilgenbusch⁹

¹Rosen Center for Advanced Computing, Purdue University, West Lafayette, IN, United States,

²Research Computing and Data, Clemson Computing & Information Technology, Clemson University, Clemson, SC, United States, ³Research Technologies Division, Office of the Vice President for Information Technology, Indiana University, Bloomington, IN, United States, ⁴Institute for Cyber-Enabled Research, Michigan State University, East Lansing, MI, United States, ⁵Department of Computational Mathematics, Science, and Engineering, Michigan State University, East Lansing, MI, United States, ⁶Department of Physics and Astronomy, Michigan State University, East Lansing, MI, United States, ⁷Facility for Rare Isotope Beams, Michigan State University, East Lansing, MI, United States, ⁸Pervasive Technology Institute, Indiana University, Bloomington, IN, United States, ⁹Research Computing, Research & Innovation Office, University of Minnesota, St. Paul, MN, United States

High-performance computing (HPC) is widely used in higher education for modeling, simulation, and AI applications. A critical piece of infrastructure with which to secure funding, attract and retain faculty, and teach students, supercomputers come with high capital and operating costs that must be considered against other competing priorities. This study applies the concepts of the production function model from economics with two thrusts: (1) to evaluate if previous research on building a model for quantifying the value of investment in research computing is generalizable to a wider set of universities, and (2) to define a model with which to capacity plan HPC investment, based on institutional production—inverting the production function. We show that the production function model does appear to generalize, showing positive institutional returns from the investment in computing resources and staff. We do, however, find that the relative relationships between model inputs and outputs vary across institutions, which can often be attributed to understandable institution-specific factors.

KEYWORDS

cyberinfrastructure, ROI, HPC, economics, value proposition, research computing and data, RCD

1 Introduction

In higher education, university leaders must balance competing priorities in which to invest funding for infrastructure. While federal funding opportunities do exist for cyberinfrastructure, it is best treated as a coordinated investment by many stakeholders, including the universities themselves.

At a given university, a vice president for research must decide if a proposed new electron microscope adds the most value, or is it a wind tunnel instead? Even within an IT organization, a CIO must choose between research IT, the enterprise resource planning system, student wireless internet, and more. In order for research computing and data to reach the top of any of these lists, a clear alignment of the value proposition to the institution's goals and measures of success must be communicated.

However, many research computing organizations are immature at communicating their value, instead approaching the importance of investing in research computing as an article of faith. Ludwig (2012) described the state of the community's quantification of costs, benefits, and their relationship as "There is not much research available to answer these questions. In fact: **almost no research**".

This claim by Ludwig is further illustrated by Broude Geva et al. (2021), where the authors noted that relatively few centers even measure the usage of their systems (41 of 69), academic outcomes (33 of 69), and ROI (32 of 69).

This study's goals are two-fold:

1. Apply the production model for cyberinfrastructure described by Smith (2024) to several additional institutions to evaluate the model's suitability for general use.
2. Based on surveyed real-world data, invert the production function to create a model for sizing cyberinfrastructure investment based on levels of output.

2 Materials and methods

2.1 Literature review

A systematic literature review of ROI from educational research grant funding can be found in Rahman et al. (2024) where the authors examine published ROI studies over a twenty year period from 2001 to 2021 and sort deliverables into output and outcomes, specifically those that are tangible and intangible. They also discuss the different types of return on investment measurement found in those studies: financial, cost benefit analysis, return on equity, and social return on investment. A case is made to define and communicate research grant output definitions early so that uniform research evaluations and measurement of the impact and ROI of research funding can be accomplished.

There are also macro-analysis studies of ROI in research infrastructures as highlighted by Del Bo (2016) in an analysis of research infrastructure investment in the European Union between 1987 and 1991. Based on prior publication the author concludes the primary method of analysis for specific projects should use a cost benefit analysis approach. The author introduces a new cost effectiveness measure that factors in publications to provide a proxy rate of return and notes the difference in definition of similar terms in macro and micro studies. The conclusion at the macro-level is a positive contribution to economic growth and productivity from investment in large research infrastructures and notes the open issues of time lag over which the activities exert an impact, in addition to spillover between geographical regions and industries.

More detailed reviews of literature on measuring the value of cyberinfrastructure that serves as the theoretical foundation for this work may be found in Smith (2024) and Stewart et al. (2022). The following sections summarize the key outcomes from these previous works.

2.1.1 ROI in cyberinfrastructure

The methods utilized in this study are described in Smith (2024), which employed the concept of the production function (Cobb and Douglas, 1928) from economics as a tool to describe how CI investment contributes to institutional output. Apon et al.'s (2010, 2015) works were some of the first studies of the ROI of cyberinfrastructure, and their exploration of how the presence of a robust cyberinfrastructure impacted institutional and departmental productivity were a key influence for Smith's model.

At Indiana University, Snapp-Childs et al. (2024) used the International Integrated Reporting Framework to assess the cost of the XSEDE cyberinfrastructure and the value of the end products created using it. Stewart et al. (2019), coining the notion of ROI_{proxy}, noted that "A value of greater than one indicates that what was done would have cost more if done by an alternative method."

Stewart et al. (2018) found positive ROI (greater than 1) for each of a campus supercomputer, Jetstream, and XSEDE by assigning financial values to them by comparing to commercial services or user surveys.

Smith (2024) developed his model to build upon Apon's, specifically adding a people dimension and addressing the decreasing utility of the Top 500 list as a proxy for an institution's computing capacity. Smith only assessed investment and outputs of one institution, however, leading to a clear opportunity for improvement by applying the production function model to more institutions than just Purdue. It is this gap that this study seeks to fill.

Several articles provide a theoretical base on selecting metrics with which to measure CI investment and institutional output. Stewart et al. (2019) described several potential outcome metrics, including the amount of grant income, new products and patents, the amount of economic impact, and new research innovations (e.g., papers, citations).

Arora and Sondhi (2024) "developed a taxonomy of the CI projects and defined a set of standard metrics for the assessment of the projects belonging to the different categories in the taxonomy." Various metric categories evaluated include outcomes like publications, citations, IP, time to degree, and the resulting MICI (Measuring the Impact of CyberInfrastructure) model introduces an impact score as a broadly applicable way to measure CI projects.

OECD (2019) proposed a "Framework for assessing the scientific and socio-economic impact of Research Infrastructures" and linked 58 potential impact indicators for research infrastructure (RI) to strategic objectives. These indicators include the number of publications, citations, high impact publications, grants, patents, students trained, and more.

In a 2014 workshop report, Berente et al. (2014) identified four dimensions for CI value outcomes: economic, science, workforce, and innovation. In a subsequent panel at SC14 (Newby et al., 2014), speakers described input investments as "physical infrastructure, energy, capital equipment and software and personnel". Academic products (outputs) are described as "Intellectual/Reputation ROI (publications, graduations, numbers of faculty and student users,

TABLE 1 Selected input and output metrics.

Input	Output
Labor: Annual salary costs	R&D expenditures reported to HERD
Capital: Total TF operated	Earned doctorates reported to SED
	Total publications produced, per Scopus
	High-impact “Nature Index” publications

and enhancement to reputation); and Financial ROI (grants and contracts, recharge fees, and patents/royalties).”

2.1.2 The production function model

Cobb and Douglas’ (1928) article “A Theory of Production” (Cobb and Douglas, 1928), described a statistical model that framed the output of production *Y*, using the inputs of capital *C*, and of labor *L*.

The production function model is represented in equation form in Equation 1.

$$Y = f(L, C) \tag{1}$$

Using a least-squares regression, Cobb and Douglas reported that the production function had coefficients of 0.75 for *L* and 0.25 for *C*, indicating a 75% and 25% proportion of labor and capital’s contribution to the output of the function.

Comparing to Cobb and Douglas’ 75:25 proportion in the academic cyberinfrastructure sector, Smith (2024) reported production function labor and capital coefficients for Purdue University of 63:31, 65:39, 60:35, or 58:36 respectively, depending on the output being modeled.

2.2 Historical dataset

A dataset was collected from five participating R1 institutions between 2011 and 2023 that included the annual year-over-year HPC capacity centrally operated and the annual university-funded salary and wage expenses relating to that centralized HPC. These self-reported, internal data are combined with publicly available metrics of outputs: R&D expenditures as reported to the NSF HERD Survey, publications, the high-impact subset of those publications tracked in the Nature Index, and earned doctorates.

Input and output metrics are summarized in Table 1, and further details about metrics are further described in Section 2.2.1 and 2.2.1.

2.2.1 Input metrics

The set of annual input and output metrics is described in further detail below:

- **TeraFLOPS of compute capacity available:** This metric represents the aggregate compute capacity centrally operated at the institution during a given year. A proxy for overall center size, this information was assembled from internal

records maintained by center leadership. These values do not include HPC resources that are funded and operated on behalf of outside entities, or lab/department scale resources.

- **Total salary costs for the HPC center:** This metric is the total salary and wage expenses for the campus Research Computing and Data (RCD) organization per internal financial systems. These dollars do not include expenditures on grant accounts, as this study is investigating *institutional* investment.

Capacity in FLOPS is used rather than a raw count of cores, nodes, or GPUs to account for the relative performance of those cores or GPUs as time goes on. For example: 10,000 7-year old cores are likely not an indicator of greater capacity or investment compared to 7,000 1-year old cores. The 1 year-old cores will most likely have a greater peak performance than the 7-year old system. Further, a center with large but extremely old resources may actually be one subject to under-investment. For labor metrics, FTE count is potentially an equivalent proxy for investment, but potentially more difficult to track historically than salary dollars, which can be reported from institutional financial systems.

2.2.2 Output metrics

The following annual data are described in further detail below:

- **HERD-reported R&D expenditures:** This metric represents the total annual institutional R&D expenditures reported to the NSF Higher Education Research and Development (HERD) survey.¹
- **Earned doctorate degrees:** This metric represents the number of Doctoral degrees awarded by each institution annually, reported to the NSF Survey of Earned Doctorates (SED).²
- **Total publications by authors affiliated with the institution:** This metric is the total number of publications reported in Scopus.³ for a given year, with an affiliation tying to the institution
- **Total publications in high-impact journals:** this metric is the total number of publications reported in Scopus for a given year, where an author with the institution’s affiliation published in one of the high-quality journals listed in the Nature Index.⁴

2.3 Current investment dataset

In Spring 2025, a survey was sent to Coalition for Academic Scientific Computation (CASC) member institutions and members of the Midwest Research Computing and Data Consortium (MWRCD) to obtain a broader dataset of system capacity and salary investment to generate a broad dataset with which to develop a benchmarking model, and create a

1 <https://www.nsf.gov/statistics/srvyherd>
2 <https://www.nsf.gov/statistics/doctorates/>
3 <https://scopus.com>
4 <https://www.natureindex.com/>

TABLE 2 Correlation coefficients—TF vs outputs, by institution.

Output metric	Institution A	Institution B	Institution C	Institution D	Institution E
HERD expenditures	0.670	0.948	0.868	0.901	0.930
Publications	0.961	0.958	0.898	0.878	0.918
Hi impact pubs	0.585	0.886	0.853	0.795	0.916
Earned doctorates	0.716	0.599	0.515	0.219	0.870

TABLE 3 Correlation coefficients—salary costs vs outputs, by institution.

Output metric	Institution A	Institution B	Institution C	Institution D	Institution E
HERD expenditures	0.295	0.895	0.706	0.857	0.767
Publications	0.543	0.876	0.647	0.786	0.751
Hi impact pubs	0.421	0.838	0.647	0.429	0.767
Earned doctorates	0.490	0.562	0.657	-0.429	0.660

base for creating a future broader-scale historical dataset, to address concerns with the small size of the historical datasets. 28 responses were received, all from R1 Carnegie Classification universities, with R&D expenditures ranging from under \$180M to \$2B.

These survey responses were merged with the most recent (2023) HERD and Survey of Earned Doctorates (SED) data for each institution, and the data were cleaned. Some respondents were unable to provide capacity in TeraFLOPS, and we worked with the respondent to calculate a value based on their reported core and GPU counts by their hardware specifications. Similarly, some respondents were unable to report a dollar amount for staff costs but provided an FTE count. In this case, the FTE count was multiplied by the median RCD professional compensation of \$90k USD reported by Maimone et al. (2023).

Each institution’s ratios of TeraFLOPS and salary costs to 2023 HERD expenditures were calculated, and the ratios then averaged to find a coefficient with which to describe system capacity and salary investment as a function of R&D expenditures.

R&D expenditures is not the only driver of computing demand. Another way to quantify the need for capacity at a research institution is in the most important product of all—its graduates. An institution that graduates high numbers of STEM PhDs will have a greater need for a well-resourced cyberinfrastructure than one that only graduates a few.

To model this additional dimension, each institution’s ratios of TeraFLOPS and salary costs to 2023 earned doctorates were calculated. As with expenditures, these relationships were then averaged to identify a coefficient with which to size system capacity and salary investment as a function of PhDs awarded.

2.4 Correlation analysis

Next, with the historical dataset, we will perform a correlation analysis of single institutional research computing investment factors with university outputs.

Hypothesis: there is universally a strong correlation between institutional inputs and outputs, as has been shown at Purdue University in Smith (2024).

2.5 Production function model

Following correlation analysis, the production function regression models will be performed first against the combined historical dataset and then individually for each institution, regressing institutional outputs on compute capacity and RCD staff investment. Regression coefficients will be summarized, compared to those for Purdue identified by Smith (2024), and the coefficients translated to real-world impact on the selected output.

Hypothesis: the combined and individual production function models will be highly-performing for all institutions examined, as with Purdue in Smith (2024).

2.6 Relative importance of coefficients

Using the lmg method implemented in the R package relaimpo (Grömping, 2007), we will calculate the proportion of the overall variance explained by each input in the model.

Hypothesis: the proportions of variance on the combined and individual institutions’ production function models will be similar to Purdue’s in Smith (2024).

2.7 Benchmark: investment as a function of outputs

Research question: Can the existing relationships between a campus’ RCD investment and its outputs be generalized to create a best practice to guide decision makers operating a university HPC center?

A question at the front of the mind of any research computing leader is “is my center appropriately resourced for the expectations

TABLE 4 Production function models—all institutions.

	Publications	Earned doctorates	HERD expenditures	High impact publications
(Intercept)	1,934.814***	301.258***	154.743***	254.473***
	(409.760)	(35.628)	(38.128)	(62.676)
TF	0.196*	0.012	0.030***	0.036*
	(0.090)	(0.008)	(0.008)	(0.014)
Salaries	1,369.031***	78.246***	144.637***	212.089***
	(189.689)	(16.493)	(17.650)	(29.014)
R ²	0.535	0.341	0.634	0.557
Adj. R ²	0.524	0.325	0.625	0.546
Num. obs.	86	86	86	86

****p* < 0.001; ***p* < 0.01; **p* < 0.05.

TABLE 5 Production function models—combined dataset.

	Publications	Earned doctorates	HERD expenditures	High impact pubs
100 TeraFLOPS	19.59	1.24	2.99	3.59
\$100k salaries	136.90	7.83	14.46	21.21
Adj. R ²	0.524	0.325	0.625	0.546

of researchers at my university?” Comparing raw numbers to each other does not provide a good benchmark—a \$300M/year research enterprise is unlikely to require the level of resources that a billion-dollar university does. Much like nations measure defense or healthcare costs as a fraction of GDP, we will measure RCD investment as a function of key outputs.

Specifically, with the benchmarking dataset we will answer these questions:

- **Benchmark Q1:** What are the observed ratios of computing capacity to outputs?
- **Benchmark Q2:** What are the observed percentage of salary costs vs. outputs?
- **Benchmark Q3:** With this information, can we present formulae for appropriately sizing an RCD center’s people and capital resources?

3 Results

3.1 Correlation analysis

First, we perform a correlation analysis using the Kendall Tau rank correlation, comparing the capital input TeraFLOPS and labor input salaries to identified outputs. Kendall correlation is used rather than Pearson due to not only to the relatively small dataset sizes per institution, but also the non-normal distribution of TeraFLOPS and salaries for each institution. For TeraFLOPS, all institutions show a strong correlation with each output, with the exception of Institution D’s earned doctorate output (Tables 2, 3).

Salary investment from all institutions shows a moderate to strong correlation with most outputs ($\simeq 0.6$ to 0.9), with Institution A’s salary correlations being the weakest overall. Institution D’s correlation with earned doctorates is not meaningful (it’s negative,

here, in fact), due to external institutional trends in doctoral degrees awarded discussed in greater detail in Section 4.1.1.

3.2 Production function model

Table 4 reports the results of the and of the production function model across the entire multi-year combined dataset ($n = 82$) of all five studied institutions. The resulting models report moderate adjusted R^2 values for each output, with statistically significant predictors.

Despite only moderately high adjusted R^2 and a limited sample size of institutions, this application of the production model to this population shows promise as first attempt to replicate Smith (2024) with a model incorporating more than a single institution. Table 5 reports the results of the combined production function model in plain terms, with, for example, each increase of 100 TeraFLOPS corresponding with an increase of \$2.99 million in annual HERD R&D Expenditures. Each \$100k investment in RCD salaries corresponds with an increase of \$14.46 million of annual HERD expenditures.

As we explore deeper, we apply the production function model to each institution individually, as performed by Smith (2024). Each subsection in Section 3.2 examines an output (publications, earned doctorates, high-impact publications, and HERD R&D expenditures) and presents each individual institution’s model results.

3.2.1 Publications

In Table 6, production function models on publications exhibit high adjusted R^2 for all institutions, 3 of which have both coefficients exhibiting a p -value < 0.05 . The high p -values for Institution D is likely due to the limited number of observations (only 8 years), and we anticipate that model predictors would be

TABLE 6 Production function models—predicting publications.

	Institution A	Institution B	Institution C	Institution D	Institution E
(Intercept)	1,508.149***	1,521.613***	4,340.957***	9,229.195***	1,987.484***
	(112.468)	(161.583)	(171.112)	(1,077.371)	(250.412)
TF	0.404***	0.082***	0.216***	0.231	0.231***
	(0.045)	(0.018)	(0.046)	(0.098)	(0.043)
Salaries	230.963*	1,135.829***	870.411***	149.197	1,340.362***
	(88.555)	(58.961)	(132.982)	(332.278)	(123.793)
R ²	0.918	0.977	0.921	0.784	0.938
Adj. R ²	0.904	0.974	0.910	0.697	0.933
Num. obs.	15	21	17	8	25

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

TABLE 7 Summary—publications output.

	Institution A	Institution B	Institution C	Institution D	Institution E
100 TeraFLOPS	40.41	8.18	21.57	23.09	23.12
\$100k salaries	0.04	113.60	87.04	14.92	134.00
Adj. R ²	0.904	0.974	0.921	0.697	0.933

statistically significant with the addition of more historical data points. Institution A's salary coefficient is low and non-significant, due to historical volatility in investment levels (Section 4.1.1).

Table 7 summarizes the results of the production function model on publications in real-world terms. For Institution C, each increase of 100 TeraFLOPs operated corresponds with an annual increase of 21.57 publications coming from Institution C, and each \$100 k investment in RCD salaries corresponds with an annual increase of 87.04 publications.

3.2.2 Earned doctorates

Table 8 describes the results of production function models on earned doctorates. For this output, these institution-level models do not exhibit the same significant relationship as found in Smith (2024).

Institution A, Institution B, and Institution C have only one significant predictor. For Institution D and Institution A in particular, the model fails to show significance, due to the external factors related to the output itself discussed in Section 4.1.1.

Table 9 summarizes the results of the production function model on earned doctorates in real-world terms. For example, we can see that for Institution E, each additional 100 TeraFLOPs aligns with with an annual increase of 2.10 PhDs awarded , and each \$100 k investment in RCD salaries corresponds with an additional 9.86 earned doctorates annually. No other institution's model provide a meaningful result.

3.2.3 HERD expenditures

For HERD R&D Expenditures, all institutions exhibit a strong production function model with a high adjusted R², indicating that using the production function as a model

appears to broadly hold true (Table 10). Only Institution A has an insignificant predictor (salaries) atypical from the other institutions, the small coefficient further reinforcing that the volatility in Institution A's investment/disinvestment cycle make the salaries variable meaningless for the purpose of modeling (Section 4.1.1).

Table 11 summarizes the results of the production function model on HERD Expenditures in real-world terms. For example, we can see that for Institution B each additional 100 TeraFLOPs represents an annual increase of \$2.13M of research expenditures, and each \$100 k investment in RCD salaries corresponds with an additional \$13.10 M of annual expenditures.

3.2.4 High-impact publications

The production function models for high-impact publications are varied—Institution E and Institution C both have models that explain significant amounts of the output, with both a high adjusted R² and significant predictors (Table 12).

As with publications (Section 3.2.1), this result with low adjusted R² and non-significant coefficients for Institution D is likely due to the limited number of observations for it in the model (only 8 years). Production function results are expected to improve with a larger set of data points. Institution A again has an insignificant salaries predictor, due to its investment/disinvestment cycles.

Table 13 summarizes the results of the production function model on high-impact publications in real-world terms. For example, we can see that for Institution E, each additional 100 TeraFLOPs represents an annual increase of 4.61 high impact, Nature Index publications, and each \$100k investment in RCD salaries corresponds with an additional 21.12 high-impact publications annually.

TABLE 8 Production function models—predicting earned doctorates.

	Institution A	Institution B	Institution C	Institution D	Institution E
(Intercept)	168.355***	308.485***	439.563***	622.061*	380.944***
	(14.787)	(24.282)	(11.909)	(202.725)	(19.239)
TF	0.026***	−0.001	−0.003	−0.020	0.021***
	(0.006)	(0.003)	(0.003)	(0.018)	(0.003)
Salaries	23.798	47.405***	38.816***	30.359	98.548***
	(11.643)	(8.860)	(9.255)	(62.524)	(9.511)
R ²	0.767	0.698	0.618	0.238	0.941
Adj. R ²	0.728	0.665	0.563	−0.067	0.936
Num. obs.	15	21	17	8	25

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

TABLE 9 Summary—earned doctorates output.

	Institution A	Institution B	Institution C	Institution D	Institution E
100 TeraFLOPS	2.62	−0.05	−0.26	−2.05	2.10
\$100k salaries	2.38	4.74	3.88	3.04	9.86
Adj. R ²	0.728	0.665	0.563	−0.067	0.936

TABLE 10 Production function models—predicting HERD expenditures.

	Institution A	Institution B	Institution C	Institution D	Institution E
(Intercept)	157.890***	122.298*	374.975***	440.168***	184.293***
	(17.901)	(46.438)	(23.840)	(58.239)	(26.299)
TF	0.046***	0.021***	0.034***	0.027**	0.032***
	(0.007)	(0.005)	(0.006)	(0.005)	(0.005)
Salaries	3.075	130.964***	112.268***	136.264***	118.495***
	(14.095)	(16.945)	(18.528)	(17.962)	(13.001)
R ²	0.817	0.906	0.922	0.986	0.937
Adj. R ²	0.787	0.895	0.911	0.981	0.931
Num. obs.	15	21	17	8	25

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

TABLE 11 Summary—HERD expenditures output.

	Institution A	Institution B	Institution C	Institution D	Institution E
100 TeraFLOPS	4.64	2.13	3.38	2.73	3.18
\$100k salaries	0.31	13.10	11.23	13.63	11.85
Adj. R ²	0.787	0.895	0.911	0.981	0.931

3.3 Relative importance of coefficients

Using the lmg method implemented in the R package relaimpo (Grömping, 2007), we can calculate the proportion of the overall variance explained by each input in the model. The “Combined” column in Table 14 shows the relative importance of each input per output, for the production function model built from the combined dataset, as well as itemized by each institution.

For example, we can see that for Institution C, Institution B, and Institution E, the production function model for HERD expenditures has similar proportions of variance explained by TeraFLOPs available on campus and salary costs, averaging 31% and 59%, respectively. These models have high adjusted R² values between 0.925 and 0.955. The impact of salaries on Institution A's models is far less than other institutions, due to its investment/disinvestment cycles described in Section 4.1.1.

TABLE 12 Production function models—predicting high-impact publications.

	Institution A	Institution B	Institution C	Institution D	Institution E
(Intercept)	195.001***	181.688***	571.074***	1,257.600**	325.057***
	(28.884)	(36.117)	(27.888)	(238.894)	(37.293)
TF	0.014	0.005	0.047***	0.009	0.054***
	(0.012)	(0.004)	(0.008)	(0.022)	(0.006)
Salaries	43.981	184.052***	134.903***	57.533	202.483***
	(22.743)	(13.179)	(21.673)	(73.679)	(18.436)
R ²	0.429	0.948	0.934	0.382	0.955
Adj. R ²	0.334	0.942	0.925	0.135	0.951
Num. obs.	15	21	17	8	25

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

TABLE 13 Summary—high-impact publications output.

	Institution A	Institution B	Institution C	Institution D	Institution E
100 TeraFLOPS	1.37	0.46	4.66	0.88	4.61
\$100k salaries	4.40	18.41	13.49	5.75	21.12
Adj. R ²	0.429	0.942	0.925	0.135	0.951

The relative importance of the combined model shows salary as approximately twice as important as TeraFLOPs of capacity as a predictor. As noted regarding [Tables 4](#) and [5](#), the relative importance for the combined dataset for the entire population is much lower than the models for each institution, due to the limited breadth and statistical noise of the historical dataset.

3.4 Benchmark models—function of R&D expenditures

Based on the 28 responses to the survey, the average ratio of TeraFLOPs of capacity to \$1M of R&D expenditures is 11.47 TeraFLOPs per million dollars. (**Benchmark Q1**), and the average percentage of salary costs vs. R&D Expenditures is 0.29% of R&D expenditures (**Benchmark Q2**).

With the value of the metrics of TeraFLOPs or salary costs *Investment* shown as a function of institutional output R&D expenditures $Investment = f(D_{herd})$, these coefficients are shown in [Equations 2](#) and [3](#) (**Benchmark Q3**).

$$TF = 11.47 \times D_{herd} \quad (2)$$

$$Salaries = 0.29\% \times D_{herd} \quad (3)$$

Using these coefficients, [Table 15](#) presents the results of the benchmarking models, as a function of R&D expenditures, based on the surveyed data in 2025. To compare these two models to reality:

By 2023, Institution E has exceeded \$845 M of expenditures, with the benchmarking model calculating 9.7 petaFLOPs and \$2.5 M of staffing for that level of expenditures. Institution E reported

10.9 of petaFLOPs of capacity, and \$2.65 M of actual salary costs to the survey as of the winter of 2024–25.

3.4.1 Estimating total budgets

[Lozado and Sahoo \(2023\)](#) for Gartner, Inc. report that “34% of enterprise IT spending is allocated to personnel salaries and benefits”. Using this formula and the known salary and wage expense at the institutions in the current investment survey dataset, we can estimate each respondent’s total annual RCD budget, and subsequently compute the fraction of R&D expenditures that budget represents.

On average, research computing budgets of the 28 institutions in the current investment survey dataset are estimated at \$5.88M, or an average of 0.79% of annual R&D expenditures.

3.5 Benchmark models—function of earned doctorates

The average ratio of TeraFLOPs of capacity to earned doctorates is 19.65 TeraFLOPs per earned doctorate. (**Benchmark Q1**), and the average salary cost per earned doctorate is \$4,696 (**Benchmark Q2**).

With the value of the metrics of TeraFLOPs or salary costs *Investment* shown as a function of institutional output earned doctorates $Investment = f(N_{phd})$, these coefficients are shown in [Equations 4](#) and [5](#) (**Benchmark Q3**).

$$TF = 19.65 \times N_{phd} \quad (4)$$

$$Salaries = \$4,696 \times N_{phd} \quad (5)$$

TABLE 14 Relative importance, all production function models.

Input metric	Inst. A	Inst. B	Inst. C	Inst. D	Inst. E	Combined
Publications						
TeraFLOPS	0.070	0.260	0.402	0.508	0.345	0.135
RCD salaries	0.211	0.716	0.520	0.276	0.593	0.400
Earned doctorates						
TeraFLOPS	0.530	0.110	0.077	0.195	0.381	0.091
RCD salaries	0.237	0.589	0.540	0.043	0.560	0.250
HERD expenditures (\$M)						
TeraFLOPS	0.724	0.340	0.435	0.450	0.420	0.154
RCD salaries	0.093	0.565	0.487	0.536	0.517	0.403
Hi Impact Publications						
TeraFLOPS	0.158	0.193	0.466	0.164	0.426	0.185
RCD salaries	0.270	0.755	0.468	0.219	0.529	0.450

Table 16 shows the estimated TeraFLOPs, salary costs, and total RCD budgets needed by an institution at various quantities of PhD graduates, based on these coefficients. To compare these two models to reality:

In 2023, Institution E awarded 810 doctoral degrees, with the benchmarking model calculating 14 petaFLOPs and \$3.8M of staffing for that number of graduates. Institution E reported 10.9 of petaFLOPs of capacity, and \$2.65M of actual salary costs to the survey, as of the winter of 2024–25.

3.6 Benchmark models—function of publications

Results of the historical dataset indicate that publications-based models are strong, therefore we present, as another option, a model publication data in the survey dataset.

The average ratio of TeraFLOPs of capacity to publication output is 1.34 TeraFLOPs per publication reported in Scopus, (**Benchmark Q1**), and the average salary cost per earned doctorate is \$3,400 (**Benchmark Q2**).

With the value of the metrics of TeraFLOPs or salary costs *Investment* shown as a function of institutional output earned doctorates $Investment = f(N_{pub})$, these coefficients are shown in **Equations 16 and 7 (Benchmark Q3)**.

$$TF = 1.34 \times N_{pub} \quad (6)$$

$$Salaries = \$3,400 \times N_{pub} \quad (7)$$

Table 17 shows the estimated TeraFLOPs, salary costs, and total RCD budgets needed by an institution at various quantities of publication output, based on these coefficients. To compare these two models to reality:

In 2023, Institution E authors had 7,649 publications in the Scopus database, with the benchmarking model calculating 10.2 petaFLOPs and \$2.61M of staffing for that number of graduates.

Institution E reported 10.9 of petaFLOPs of capacity, and \$2.65M of actual salary costs to the survey, as of the winter of 2024–25.

4 Discussion

With these results, we present a set of metrics and models that will allow university leaders to generalize the work by Smith (2024) and validate this initial model for institutions beyond Purdue University, and apply the concept of supercomputing as a function of institutional outputs to capacity planning and budgeting.

Although the historical dataset has a limited sample size (5 institutions, 8–24 years per institution), this study provides some confirmation that, with some exceptions such as earned doctorates, investment in either computing resources or staffing corresponds with increases of multiple institutional outputs. Investment in the RCD center's staff contributes for the largest relative importance in all models, as with Smith. Data from many more institutions, and over a longer time horizon, must be collected and analyzed in order to provide more conclusive results that fully reflect the general trends.

Specific models based on the current investment survey dataset are provided to give administrators a tool with which to size their investment based on the institution's outputs. On average, the 28 institutions in the current investment survey dataset operate 11.47 TeraFLOPs of central capacity per \$1 M of R&D Expenditures, and spend an amount equal to 0.29% of their R&D expenditures on research computing salaries.

4.1 Production function model results

Tables 18, 19 summarizes the results of production function models from the historical dataset for all outputs, further separated by institution and for the entire dataset.

For the production model across the entire population, however, the resulting models report moderately strong adjusted R^2 values for each output, with statistically significant predictors.

TABLE 15 Center investment benchmarks—model results based on TeraFLOPS.

2023 R&D Exp (\$M)	Modeled TF capacity (2025)	2023 R&D Exp (\$M)	Modeled RCD salaries (in 2025 \$M)	Modeled total budget (in 2025 \$M)
\$1,900	21,784	\$1,900	\$5.59	\$16.45
\$1,500	17,198	\$1,500	\$4.42	\$12.99
\$1,200	13,758	\$1,200	\$3.53	\$10.39
\$1,000	11,465	\$1,000	\$2.94	\$8.66
\$850	9,745	\$850	\$2.50	\$7.36
\$750	8,599	\$750	\$2.21	\$6.49
\$400	4,586	\$400	\$1.18	\$3.46
\$200	2,293	\$200	\$0.59	\$1.73

TABLE 16 Center investment benchmarks—model results based on earned doctorates.

2023 earned doctorates	Modeled TF (2025)	2023 earned doctorates	Modeled RCD salaries (in 2025 \$M)	Modeled total budget (in 2025 \$M)
800	15,718	800	\$3.76	\$11.05
700	13,753	700	\$3.29	\$9.67
600	11,789	600	\$2.82	\$8.29
500	9,824	500	\$2.35	\$6.91
400	7,859	400	\$1.88	\$5.52
200	3,929	200	\$0.94	\$2.76

TABLE 17 Center investment benchmarks—model results based on publications.

2023 publications	Modeled TF (2025)	2023 publications	Modeled RCD salaries (in 2025 \$M)	Modeled total budget (in 2025 \$M)
20,000	26,744	20,000	\$6.82	\$20.07
14,000	18,721	14,000	\$4.78	\$14.05
10,000	13,372	10,000	\$3.41	\$10.04
6,000	8,023	6,000	\$2.04	\$6.02
3,000	4,011	3,000	\$1.02	\$3.01
1,000	1,337	1,000	\$0.34	\$1.00

Further discussion of this combined model can be found in Section 3.2.

Section 3.3 describes the relative importance of each predictor to each output. The relative importance of the combined model shows salary as approximately two to three times as important as TeraFLOPs of capacity as a predictor (Table 14).

Most output metrics show still strong (> 50%) average relative importance of their capital and labor predictors (as in Smith, 2024), at 60%/30%. Table 14 further reports the specific relative importance of each predictor to each output for each individual institution.

4.1.1 Local factors influencing quality of models

External factors with Institution A's and Institution D's salary investment and earned doctorates, respectively, limit the quality and utility of several models.

Observed both in correlation analysis and several production function models, the weakness of earned doctorates in Institution D's models is likely due to a fluctuating trend in doctoral degrees awarded there (Figure 1), rising in the early 2000s, followed by a highly variable period over the last decade, generally declining.

Furthermore, the overall weak results predicting earned doctorates at multiple institutions may indicate that Smith (2024) studying Purdue may in fact be for an atypical institution in this set. Purdue's output in earned doctorates has steadily climbed since 2003, rising to the 3rd largest producer of PhDs in the 2023 Survey of Earned Doctorates, with only one other institution in the historical dataset ranking in the top 25. If data were available, performing this analysis for other very large producers of PhDs such as Michigan, Stanford, or Texas A&M to reproduce the results from Smith would be informative.

Section 4.3.1 describes future enhancements to the models that will mitigate effects such as these.

While still positive, Institution A shows weaker correlations than other institutions due to variability in its RCD salary investment over time (Figure 1). This variability is attributed to shifting institutional priorities. RCD salary investment grew steadily until the 2017 retirement of the Provost, the President, and a research-focused CIO. This significant change in leadership was followed by a period of steady budget reductions until the current CIO, hired in 2021, reversed the decline. This external factor suggests that salary investment for Institution A may be a weaker predictor than was observed in previous work, or with other institutions in this study.

TABLE 18 Historical dataset summary—vs. **Smith (2024)**.

Publications	Historical dataset		(Smith, 2024)	
	Coefficients	Rel. Imp.	Coefficients	Rel. Imp.
100 TeraFLOPS	19.59	14%	7.14	29%
\$100k RCD salaries	136.90	40%	155.30	65%
Earned doctorates				
100 TeraFLOPS	1.24	9%	2.55	31%
\$100k RCD salaries	7.83	25%	7.36	42%
HERD expenditures (\$M)				
100 TeraFLOPS	2.99	15%	2.59	25%
\$100k RCD salaries	14.46	40%	9.04	43%
High impact publications				
100 TeraFLOPS	3.59	19%	3.34	35%
\$100k RCD salaries	21.21	45%	22.98	61%

TABLE 19 Model results—all production function models.

Input metric	Inst. A	Inst. B	Inst. C	Inst. D	Inst. E	Combined
Publications						
100 TeraFLOPS	40.41	8.18	21.57	23.09	23.12	19.59
\$100k Salaries	0.04	113.60	87.04	14.92	136.9	136.50
Adj. R ²	0.904	0.974	0.921	0.697	0.933	0.524
Earned doctorates						
TeraFLOPS	2.62	-0.05	-0.26	-2.05	2.10	1.24
Salaries	2.38	4.74	3.88	3.04	9.86	7.83
Adj. R ²	0.728	0.665	0.563	-0.067	0.936	0.325
HERD Expenditures (\$M)						
100 TeraFLOPS	4.64	2.13	3.38	2.73	3.18	2.99
\$100k salaries	0.31	13.10	11.23	13.63	11.85	14.46
Adj. R ²	0.787	0.895	0.911	0.981	0.931	0.625
Hi Impact Publications						
100 TeraFLOPS	1.37	0.46	4.66	0.88	5.37	3.59
\$100k salaries	4.40	18.41	13.49	5.75	20.25	21.21
Adj. R ²	0.429	0.942	0.925	0.135	0.951	0.556

4.2 Applications of benchmarking as a function of output

With outputs and the coefficients in the equations $Investment = f(D_{output})$ identified, there are several situations where a university leader might apply them.

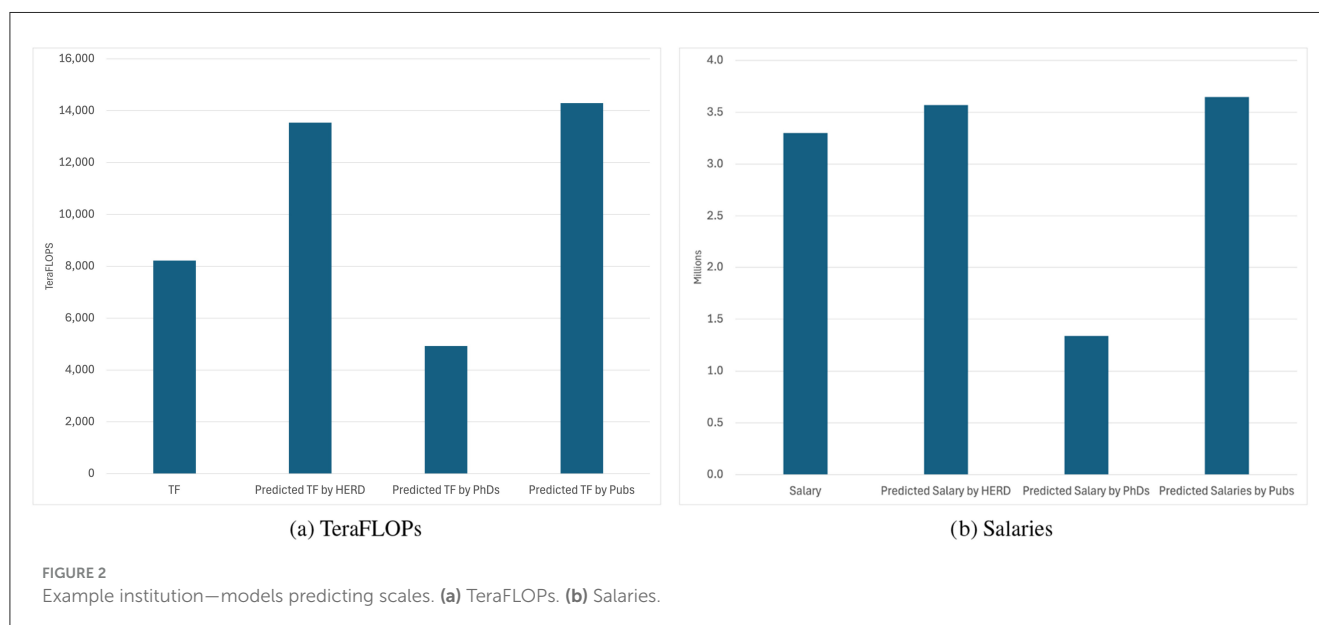
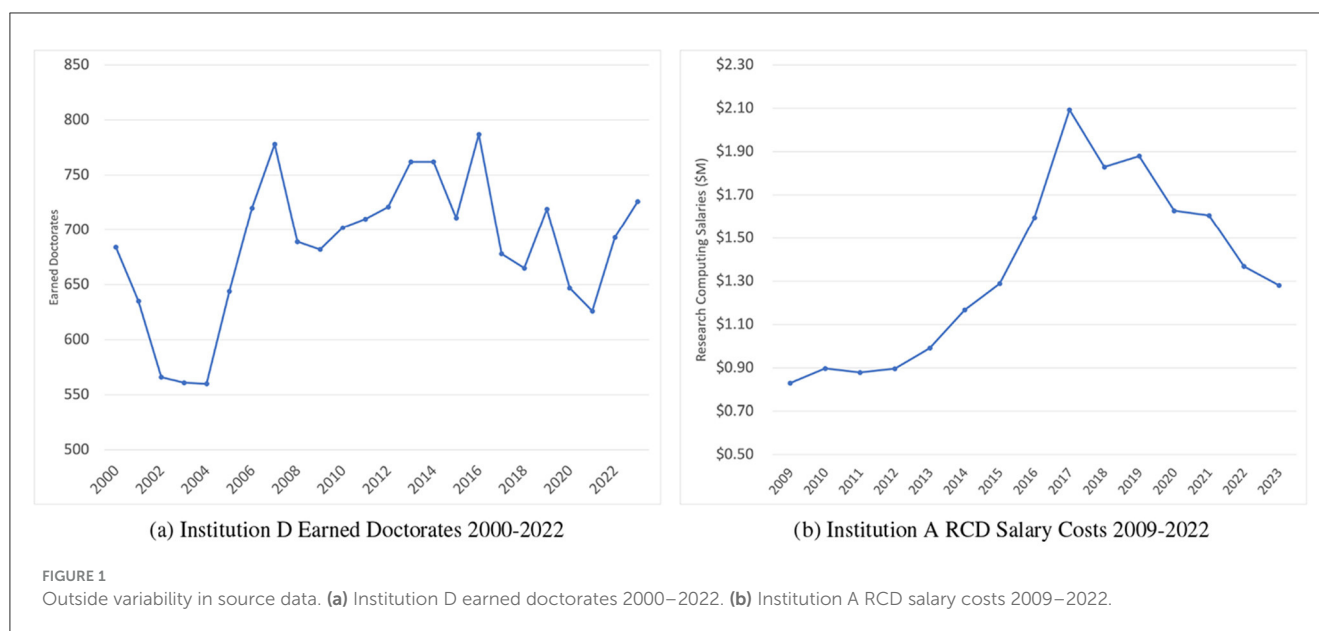
4.2.1 Selecting an output as a driver of decision-making

Benchmarking models based on R&D expenditures, earned doctorates, or publications present slightly different pictures for the same institution.

Local factors certainly do come into play: consider universities with very large research expenditures but relatively low numbers of earned doctorates (Columbia), or ones with extremely large numbers of each (Michigan). The ideal value for a given university may lie at a value in between results provided by each model.

Figures 2a, b show modeled vs actual data for an example survey respondent, an R1 from the central United States. In the case of this institution, their actual capacity in TeraFLOPS is almost right in between the values predicted by expenditures and earned doctorates, while their salary investment aligns much closer to the model based on HERD expenditures.

In both examples, the results from the model based on publications is very near to the one based on R&D expenditures. In the historical dataset, publications and R&D expenditures are



highly correlated ($\tau_{pub} = 0.985, p < 0.05$). R&D expenditures are a leading indicator of future publications and may be a better practical metric to drive decision-making.

4.2.2 Institutional benchmarking

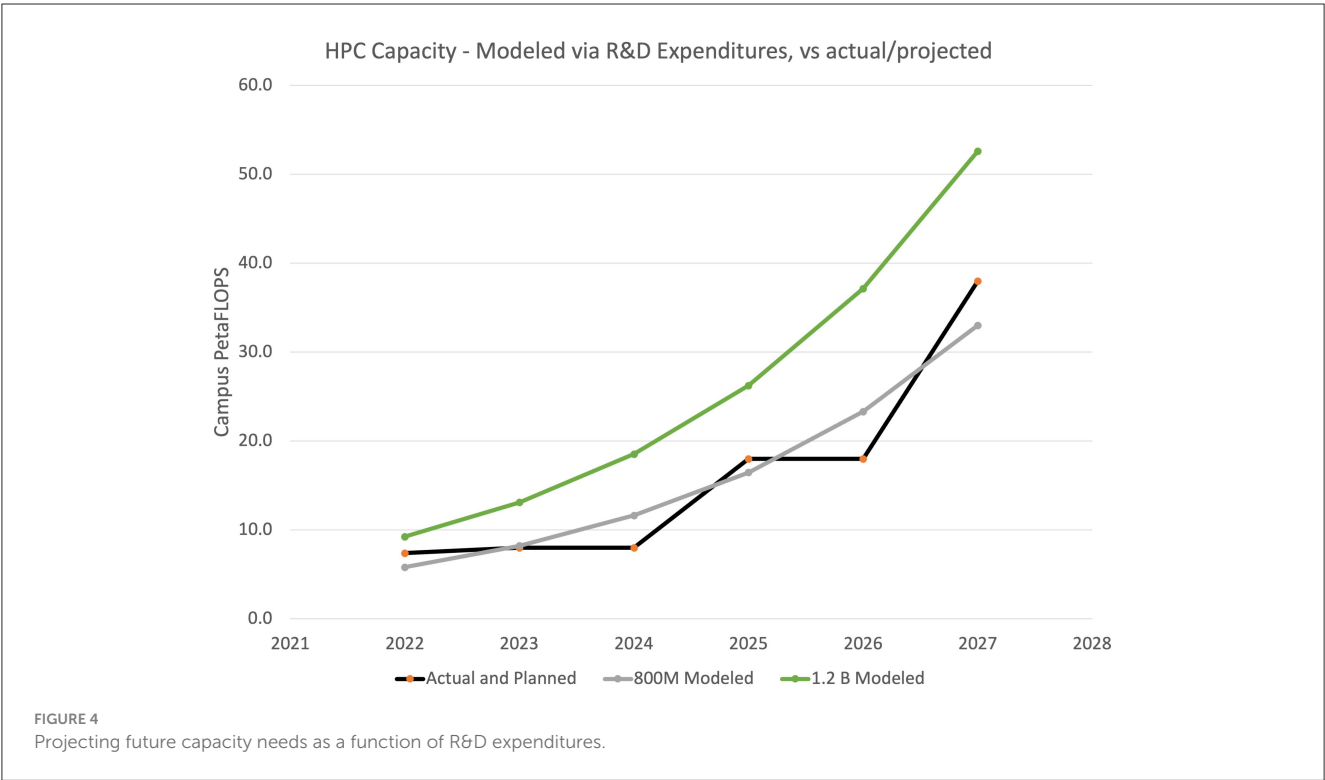
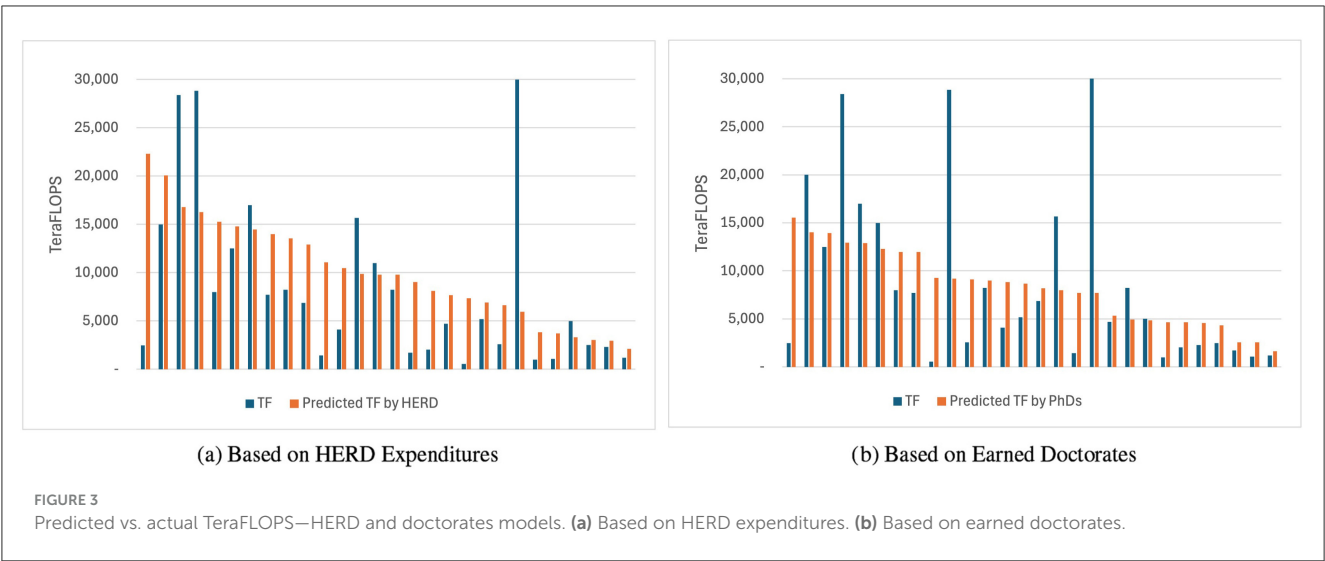
University administrators have finite resources and may ask themselves, “what size of an HPC operation does my institution need?” Under-resourcing risks impacting competitiveness, and over-resourcing is wasteful. Although the model, with only 28 respondents, may not be sufficiently fine-grained in the middle for subtle adjustments, it may already be useful at either end of the spectrum. **Figures 3a, b** show actual vs predicted TeraFLOPs for all 28 respondents, predicted by HERD expenditures and by earned doctorates. In these figures we see

many institutions’ actual capacity near their predicted value. Some outliers exist, however—both high and low—suggesting that some institutions may not be correctly resourced relative to their output.

4.2.3 Future capacity planning

The 8 to 24 years of historical data used to inform the benchmarking models in Section 2.2 provide us with an observation of historical trends across these five institutions, specifically showing an annual growth trend of 41% in operated TeraFLOPs, rather closely tracking Moore’s Law improvements.

With this year-over-year growth, **Figure 4** shows future modeled capacity at \$800 M and \$1.2 B expenditure levels, common output levels for many major universities. **Figure 4** also shows the



real deployed capacity at one university, including the early 2025 deployment of a new 10 PF system, and projections for future deployments based on long-term budget planning.

A university leader may use this model—or one based on future earned doctorates—to plan future investments. For example, if this university were to be a billion dollar research enterprise in 2028, the model suggests that it may need 43–61 PetaFLOPS of capacity.

In the real world, a \$1.2 B university deploying a 50 PetaFLOP system in 2025 may look like an extremely large investment today, but looking ahead to 2027, such a system places them right in line with their expected 2027 point on the curve, in advance.

4.3 Future work

4.3.1 Production function model

As discussed in Section 4.1.1, the external factors of declining and shifting trends in the doctoral degrees awarded at Institution D complicated modeling the earned doctorate output for that institution. Future work may explore a hypothesis that a university’s largest fields in terms of PhDs awarded are likely the CI users, and refine the earned doctorate models to use “earned doctorates from the institution’s top X% of fields of science” as an output instead of the entire number of PhDs awarded.

Similarly, the boom-and-bust investment/disinvestment cycles at Institution A complicate attempts to model the relationship between investment and output. While a counter-example of an institution with a mixed track record for maintaining its investment in research computing is useful to have, as part of a small population its presence makes model outputs in this study inconclusive at best.

Finally, it stands to reason that there is a point of diminishing returns for institutional investment in one of the factors examined, with there being a point where an even bigger supercomputer stops adding additional value. Future work could potentially build on the interaction analysis touched upon by Smith (2022) and analyze the interactions between people and capital investments. This would address questions that can more precisely guide institutional decision-making—for example, it could answer the question of whether the rate of impact on HERD expenditures from investing more in computational resources is different at higher or lower levels of investment in RCD center staff.

4.3.2 Benchmarking model

First, a benchmarking and scaling model may have a complex set of factors than simply R&D expenditures or earned doctorates. Future work may identify a better-tuned model potentially combining both of these factors, or adding new ones such as undergraduate populations, the mix of research at an institution, public/private status, among others.

Next, the rapid growth in AI compute will require more nuance in the capacity metric used herein—mixed or lower-precision FLOPS will soon far outpace FP64 at university centers, and the model will need to evolve to reflect that.

Further exploration and breakdowns of institutional characteristics can be incorporated, such as the presence of a medical school or status as a land grant institution.

Finally, in the future, we aim to continue to perform this survey annually, with the goal of adding institutions and building a historical corpus of data with which to establish longer trends. This ongoing data collection would create a dataset suitable for statistically powerful production function models across many institutions.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.nsf.gov/statistics/srvyherd>,

<https://www.nsf.gov/statistics/doctorates/>, <https://scopus.com>, <https://www.natureindex.com/>.

Author contributions

PS: Conceptualization, Data curation, Investigation, Methodology, Writing – original draft, Writing – review & editing. JG: Data curation, Writing – review & editing. DH: Conceptualization, Data curation, Writing – review & editing. BO'S: Conceptualization, Data curation, Writing – review & editing. WS-C: Conceptualization, Data curation, Methodology, Writing – review & editing. JW: Conceptualization, Data curation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This material was based upon work supported by the National Science Foundation under award numbers 2005632 and 2321090 (Smith), 2227627 (Snapp-Childs), and 1445604 and 2005506 (Hancock).

Acknowledgments

Elements of this work are based on the doctoral dissertation by Smith (2022).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Apon, A., Ahalt, S., Dantuluri, V., Gurdgiev, C., Limayem, M., Ngo, L., and Stealey, M. (2010). High performance computing instrumentation and research productivity in U.S. universities. *J. Inform. Technol. Impact* 10, 87–98. Available online at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1679248
- Apon, A. W., Ngo, L. B., Payne, M. E., and Wilson, P. W. (2015). Assessing the effect of high performance computing capabilities on academic research output. *Econ. 48*, 283–312. doi: 10.1007/s00181-014-0833-7
- Arora, R., and Sondhi, S. (2024). Towards developing an open-infrastructure for assessing the progress, success, and impacts of cyberinfrastructure projects. *SN Comp. Sci.* 5:617. doi: 10.1007/s42979-024-02961-8
- Berente, N., Howison, J., King, J. L., Cutcher-Gershenfeld, J., and Pennington, R. (2014). Leading cyberinfrastructure enterprise: value propositions, stakeholders, and measurement. *SSRN Elect. J.* doi: 10.2139/ssrn.2416247
- Broude Geva, S., Chalker, A., Hillegas, C., Petravick, D., Sill, A., and Stewart, C. (2021). "Results from a second longitudinal survey of academic research computing and data center usage: expenditures, utilization patterns, and approaches to return on investment," in *Practice and Experience in Advanced Research Computing, PEARC '21* (New York, NY: Association for Computing Machinery).
- Cobb, C. W., and Douglas, P. H. (1928). A theory of production. *Am. Econ. Rev.* 18, 139–165.

- Del Bo, C. F. (2016). The rate of return to investment in r&d: The case of research infrastructures. *Technol. Forecast. Social Change* 112, 26–37. doi: 10.1016/j.techfore.2016.02.018
- Grömping, U. (2007). Relative importance for linear regression in R: the package relaimpo. *J. Statist. Softw.* 17, 1–27. doi: 10.18637/jss.v017.i01
- Lozado, C., and Sahoo, S. (2023). *Optimize your IT Staffing Levels in 3 Steps*. Arlington, VA: Gartner, Inc.
- Ludwig, T. (2012). “The costs of HPC-based science in the exascale era,” in *Proceedings - 2012 SC Companion: High Performance Computing, Networking Storage and Analysis, SCC 2012* (Salt Lake City, UT: IEEE), 2120–2188.
- Maimone, C., Brown, C., Grash, K., Reidy, C., and Stauffer, A. (2023). “Compensation of academic research computing and data professionals,” in *Practice and Experience in Advanced Research Computing 2023: Computing for the Common Good, PEARC '23* (New York, NY: Association for Computing Machinery), 145–152.
- Newby, G., Apon, A., Berente, N., Eigenmann, R., Fratkin, S., Lifka, D., and Stewart, C. (2014). *Return on Investment from Academic Supercomputing: SC14 Panel*. New Orleans, LA: IEEE. Available online at: <https://scholarworks.iu.edu/iusrrest/api/core/bitstreams/09b0e7fc-34d7-4e6d-ba7c-fc432e5cc0d7/content>
- OECD (2019). *Reference Framework for Assessing the Scientific and Socio-Economic Impact of Research Infrastructures*. OECD Science, Technology and Industry Policy Papers, No. 65. Paris: OECD Publishing. doi: 10.1787/3ffee43b-en
- Rahman, A. T. A., Hanafiah, M. H., Abd Hamid, U. M., Ghani, N. A. M., and Zain, N. A. M. (2024). Return on investment from educational research grant funding: deliverables and measurement. *Int. J. Eval. Res. Educ.* 2252:8822. doi: 10.11591/ijere.v13i1.26632
- Smith, P. (2022). *The Value Proposition of Campus High-Performance Computing Facilities to Institutional Productivity - A Production Function Model*. West Lafayette, IN.
- Smith, P. (2024). The value proposition of campus high performance computing facilities to institutional productivity: a production function model. *SN Comp. Sci.* 5:551. doi: 10.1007/s42979-024-02888-0
- Snapp-Childs, W. G., Hart, D. L., Costa, C. M., Wernert, J. A., Jankowski, H. E., Towns, J. W., et al. (2024). Evaluating return on investment for cyberinfrastructure using the international integrated reporting <ir> framework. *SN Comp. Sci.* 5:558. doi: 10.1007/s42979-024-02889-z
- Stewart, C., Wernert, J., Costa, C., Hancock, D. Y., Knepper, R., and Snapp-Childs, W. (2022). “Return on investment in research cyberinfrastructure: State of the art,” in *Practice and Experience in Advanced Research Computing, PEARC '22* (New York, NY: Association for Computing Machinery).
- Stewart, C. A., Apon, A., Payne, R., Slavin, S. D., Hancock, D. Y., Wernert, J., et al. (2019). “Assessment of financial returns on investments in cyberinfrastructure facilities,” in *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning) - PEARC '19* (New York, New York: ACM Press), 1–8.
- Stewart, C. A., Hancock, D. Y., Wernert, J., Link, M. R., Wilkins-Diehr, N., Miller, T., et al. (2018). *Return on Investment for Three Cyberinfrastructure Facilities: a Local Campus Supercomputer; the NSF-Funded Jetstream Cloud System; and XSEDE (the eXtreme Science and Engineering Discovery Environment)* (Zurich: IEEE). doi: 10.1109/UCC.2018.00031