The Development of an End to End Life Cycle Assessment Framework for Energy and Carbon Footprint Associated with Globally Distributed Data Center Services

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Design

in School of Architecture

Eric Kumar

Carnegie Mellon University, Pittsburgh, PA

B.S., Mechanical Engineering, Sacramento State University
M.S., Fire Protection Engineering Engineering, Cal Poly San Luis Obispo

Doctoral Committee:

Dr. Erica Cochran Hameen Ph.D - *Committee Chair*

Assistant Professor Carnegie Mellon University, School or Architecture

Dr. Gwen DiPietro Ph.D

Carnegie Mellon University, Civil & Environmental Engineering

Dr. Arman Shehabi Ph.D

Research Scientist & Deputy Group Leader Sustainable Energy Systems Group Energy Analysis & Environmental Impacts Division Lawrence Berkeley National Laboratory

Dr. Inês Azevedo Ph.D

Associate Professor of Energy Resources Engineering, Stanford University & Senior Fellow at the Woods Institute for the Environment

September, 2020

©2020 Eric Kumar. *Some rights reserved.* Except where indicated, this work is licensed under a Creative Commons Attribution 3.0 United States License. Please see `http://creativecommons.org/licenses/by/3.0/us/` for details.

The views and conclusions contained in this document are those of the author, and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, current or past employers, the U.S. government, or any other entity.

Keywords: data centers, building energy models, network traffic, sustainable engineering, green design.

Abstract

In this dissertation, I demonstrate a comprehensive data center energy and carbon footprint model that consists of four parts. First, I develop a simulator that characterises internet traffic to five internationally dispersed data centers for seven real world internet services. I implement the simulation to indicate the workload demand from each service to each data center. Second, I take the data center wise workloads and construct them with a physics based building energy modeling tool, EnergyPlus, as the coefficients to the IT load profiles that each data center must support every hour of a year. Third, I take the results from the EnergyPlus model and quantify the marginal costs of the energy in terms of Carbon footprint; where the respective region's power grid consists of a mix of renewable and carbon based energy sources. Fourth, I couple the marginal energy costs model with a model that quantifies the environmental costs embodied in the materials that the data center is composed of. Through this dissertation I contribute a comprehensive model that quantifies the end to end carbon footprint of global data center systems.

This work is dedicated to

... my daughter, to whom it shows that anything is possible given the dedication.

... my parents, who left paradise to give me an opportunity that my ancestors never had.

... my wife, who has shared my sacrifices and paid the opportunity costs along my path.

Acknowledgments

TBD

Contents

List of Figures

List of Tables

0.1 Introduction

Data centers play a critical role in shaping modern economies and the lives of people around the world. They house information technology (IT) hardware and store data for financial institutions, social media accounts, and virtual meeting forums amongst many other things. In their physical embodiment, networks of data centers are global scale systems. Within these systems, each data center's size can be as large as a college campus, and consume 100 times the amount of power every hour as residential facilities do in a day. Moreover, compared to residential facilities, data centers are subject to bulks of accelerated turnovers in technology. This technology turnover results in the housed IT equipment having a useful life that only spans two to three years. The short useful life of equipment increases the ecological footprint of these systems, due to all the materials that traverse through it over its lifetime. Therefore, it is extremely important to develop a comprehensive global level modeling framework to assess data center ecological footprints.

Today there are several metrics, models, and methods available that allow us to assess energy use of data center systems. At the building scale, the most prevalent metric is the data center power usage effectiveness (PUE). PUE couples the IT loads with the building and indicates the relative efficiency of the building utilities compared to IT workloads. For PUE calculations, the IT workloads are represented in terms of power demand of the IT equipment.

However, such an approach neglects the power inefficiencies associated with hardware and the sustainability of the source energy powering the data center. An exception to just energy use modeling is the CLEER online tool. CLEER provides data center level models that account for all the materials that data center systems are composed of. Yet, CLEER's publicly accessible capability does not allow for modeling of dynamic workloads that are the characteristic of internet services, nor does it support explicit building energy models. A footprint accounting for all the carbon embodied in materials and in the operational energy will indicate the global warming potential attributed to the respective data center system.

To fill the gaps noted above, I demonstrate a comprehensive data center energy and carbon footprint modeling framework which is discussed in the four chapters of this dissertation. In the first chapter, I develop a simulator that characterises internet traffic to five internationally dispersed data centers for seven real world internet services. I implement the simulation to indicate the workload demand from each service to each data center. In the second chapter, I take the data center workloads and construct them with a physics based building energy modeling tool, EnergyPlus. In EnergyPlus the traffic is used as coefficients to the IT load profiles that each data center must support for every coincident hour of a year. Then in the third chanter, I take the results from the EnergyPlus model and quantify the marginal costs of the energy in terms of carbon footprint; where the respective data center's regional power grid consists of a mix of renewable and carbon based energy sources. Finally in the fourth chapter, I couple the marginal energy costs model with an environmentally extended economic input-output model that quantifies the environmental costs embodied in the materials that the data center is composed of.

Through this dissertation I contribute a comprehensive model that quantifies the end to end carbon footprint of global data center systems. Researchers and data center operators can use my demonstrated framework for evaluating current data center global warming potential or to make strategic decisions on how to deploy future data centers.

Chapter 1

Internet Traffic Profile

1.1 Introduction

In this chapter

Chapter 2

Building Energy Model of Data Centers

Chapter 3

Marginal Cost of Energy for Data Centers

Chapter 4

Life-Cycle Energy and Carbon Footprint Modeling with Data Center Building Energy Models

4.1 Introduction

In this chapter, the aim is to quantify the end to end life cycle costs of data centers by extending the operational models developed in the previous chapters. Those operational models have provided an indication of system level energy for a network of data centers and their marginal carbon dioxide footprints. Although, the presented models are a good proxy for the environmental costs of data center operations, they don't account for the energy and carbon footprint embodied in all the materials that the data centers are composed of.

To assess the end to end environmental impact of data centers, this chapter describes a three-step hybrid life-cycle analysis inclusive of operations and embodied inventories of the physical data center infrastructure. First, a review of the previously developed models for building energy and marginal costs of energy generation is provided. These two models together provide an indication of the energy and carbon footprint during the operational phase of a data center's life. Second, a life cycle modeling framework using an economic input-output (EIO) analysis model extended to environmental costs is introduced. The inputs to the EIO are constructed in this chapter based on literature reviews and the researcher's industrial experience. The third and final step provides a global scale end-to-end assessment for the energy and carbon footprint of each data center language pair analyzed in the previous chapters. With the global view, the global environmental costs of discrete language (service) can be assessed.

4.1.1 Motivations for an end to end life cycle cost model

In terms of scale, US data centers consumed 700 billion kWh in 2016. That was 1.8% of the total electricity produced in the country according to a United States Department of Energy (DOE) report [4]. To drive further intuition of their scale, a typical 100-MW data center at peak load consumes the same amount of energy in an hour as 100 homes do in a month. Given a data centers power demands, it is not surprising that operational energy use has a high sensitivity towards their total cost of ownership (TCO); making it a key metric in TCO based design decisions. While optimization for use phase energy may significantly reduce the carbon footprint of a data center (given the source energy mix does not change), it does not account for other phases in the data center's end to end life cycle. Inventories from embodied life cycle phases such materialization, transit, maintenance, and end of life are left out of the operational phase energy models that have become prevalent indicator of data center sustainability.

There are two predominant paradigms for evaluating the embodied environmental inventories for any product that has been altered by technology (techno-sphere). The first method is process based. The complexity of a process-based model is greatly influenced by the boundary conditions of the study. If the boundary is demarcated between the biosphere and the techno-sphere, then the number of distinct life cycle processes to quantify explodes by 500 times for a simple pen [5]. The alternate method is based on Leontief's macro-economic proxies that exploits economic correlations between industrial sectors. In macro-economic models, a matrix with rows and columns equal to the number of sectors in the economy is populated with the cost relationship between the row-sector and the corresponding column-sector. Industrial-sector macro-economic proxies reduce the problem space significantly as only one cost vector is required as input to analyze an entire economy. This research combines the two paradigms and presents a hybrid life-cycle assessment model of data centers that can be used to support data center design decisions.

The structure of the paper is as follows. First, some technical background about data centers is provided along with a synthesis of similar works. Then in the methodology section a dynamic model to quantify the operational and embodied costs of data center infrastructure is described. The results from the methods as then presented in the results section. This article concludes by summarizing its findings and suggesting the future direction in this area of research.

4.2 Background

4.2.1 Technical Overview of Data Centers

Modern internet data centers are district scale systems, spanning campuses that are hundreds of acres. They may contain several hundred thousand pieces of information technology (IT) equipment. IT equipment consists of physical servers, network hardware nodes, and digital storage media. Theses pieces of IT equipment sit alongside the data center's district scale cooling and electrical distribution plants housed in warehouse-scale built environments. The environmental footprint of a data center spans the full breadth of these physical pieces of infrastructure.

At their scale data centers receive power through medium voltage connections with the local utility's grid. The alternating-current voltage may then be stepped down in several steps, but ultimately, it is rectified to be used by the sensitive electronic components in the information technology hardware. Each step-down is a point of power inefficiency, with the alternating-current to direct-current conversion being the biggest point of power loss. Furthermore, anywhere that the step down or conversion occurs inside the building, the electrical inefficiencies are manifested as heat.

The heat from the electrical inefficiencies, along with the heat emitted by the IT equipment transistor state transitions and their current leakage, must be rejected to the outside of the building space by mechanical means. At a fundamental level, a mechanically driven fluid mover is needed to convey the heat from indoors to outdoors or another reservoir. In single pass cooling systems fans intake outside-air and force them through the IT equipment, capturing any heat and carrying it outside of the buildings. More complex cooling systems may include liquid or gas refrigerant medium thermodynamic cycles between the buildings and reservoirs, with the refrigerant medium capturing heat at either the building room level or at the scale of IT equipment. Precise modeling of such data center building systems with intense therm-power dynamics is now a manageable task in building energy modeling software [6, 7].

The embodied costs of IT equipment and building systems yield additional environmental costs for a data center's life cycle. The rate of innovation for IT equipment and the ever-increasing demands from the software applications creates a capital market where TCO of one generation of IT hardware rapidly increases relative to newer IT hardware solutions. The capital of cost/performance trade-off makes the positive TCO life of IT equipment between two to five years as observed by Shehabi in the DOE report [4]. This relatively short lifetime of IT further compounds the embodied costs of data centers. For example, through a 20-year data center building life, four to ten generations of IT transit through the facility. Disparities in lifetimes and dimensional scale differences between buildings and microchips make data center embodied inventory modeling complex. However, recently hybrid life cycle assessment models have shown to be effective in quantifying the embodied costs of data centers [5, 8].

Information technology equipment requires some further insights in order to make its impact to data center life cycle analysis more concrete and directed in scope. At the heart of information technology equipment are microprocessor chips. Modern chips are composed of billions of transistors which have been getting smaller in size since their first applications in electronics signal processing in the 1960's. Transistors have been the key enabler for the compaction and power efficiency gains of electronic devices over the years and they've also been shown to have one of the most dominant environmental costs within electronic products [9, 10].

Prior to the mid-2010's, transistors were composed of planar or 2-D architectures. 2-D transistors inherently had limited operational power efficiencies due to higher voltage and current leakage compared to the novel 3-D processors in the market today. Specifically it's the 3-D processes that have allowed significant operational efficiency gains for data centered operators, yet studies assessing the 3-D transistor architecture's impact to data center life cycle costs are lacking. The 2-D to 3-D transition is a recent example of rapid rate of adaption for information technology equipment that make generalized environmental impacts studies for transducer technology obsolete in two to three years [11]. The frequent churn of technology also drives rapid changes in the manufacturing process of the chips. These rapid changes in semiconductor-manufacturing processes necessitates a parametrically scalable framework where transistor chips can be evaluated in isolation from other server components.

4.2.2 Similar Works

In this section similar works that have quantified the environmental foot print of data centers are presented. Data center life cycle assessment works come from industrial operators [5], academia [8, 12], federal agencies [13], and industrial consortium's [14]. Two of the reviewed works are conducive to replicate from the ground up [5, 8], while another serves as a guideline [14], and another provides an online interactive tool to assess the footprint of targeted classes of Cloud services [13].

From the operators perspective Shah, demonstrates an end to end life cycle assessment of data center systems [5]. Shah uses a hybrid model inclusive of process based and economic input/output assessment frameworks to assess a single a data center, while using a static model for use-phase power. Whitehead, extends Shah's hybrid work and demonstrates the life-cycle costs of a real data center and sets an explicit functional unit of 1-kW of provisioned capacity [8].

The recent focus on product operational energy efficiency motivated Kline's study of the trade-offs between operational energy and the embodied costs of information technology equipment [12]. Although their bases for the embodied costs are process based, their literature does not provide sufficient insight for others to reproduce the work. Similarly, The Green Grid's Data Center Life Cycle costs guidelines outline the end goal of a data center life cycle analysis. It classifies several key attributes that need to be considered, but lacks references to explicit procedures that must be followed to achieve the goals.

The Cloud Energy and Emissions Research (CLEER) Model provides a browser based user interface to compare the environmental costs of on-premise server based services with hypothetical cloud-based systems that would provide an equivalent service [13]. CLEER's analysis is transparent and inclusive of embodied and operational costs, but it does not dynamically couple the embodied or operational costs into the model. The presented set of past works has inspired the methodology of this research as described in the next section.

4.3 Methodology

4.3.1 Functional Unit and Reference Flows

Life cycle assessment studies require a functional unit of performance of the system under study for use as a normalized reference point. As a reference point for data centers, there is an industry wide consensus that power is the best indicator for a data center's workload capacity [5, 8, 1]. Based on the consensus, the functional unit adapted for this research is chosen to be 1-kW of provisioned power per year.

Furthermore, for the abstracted functional unit to be used in comparisons between different data center design scenarios requires it to be translated to reference flow values. In this work the reference flow is constrained to 1-year of data center operations with the globally provisioned power footprint of each of the language abstracted internet services. The language (internet service) to data center distribution is indicated in Figure 4.1.

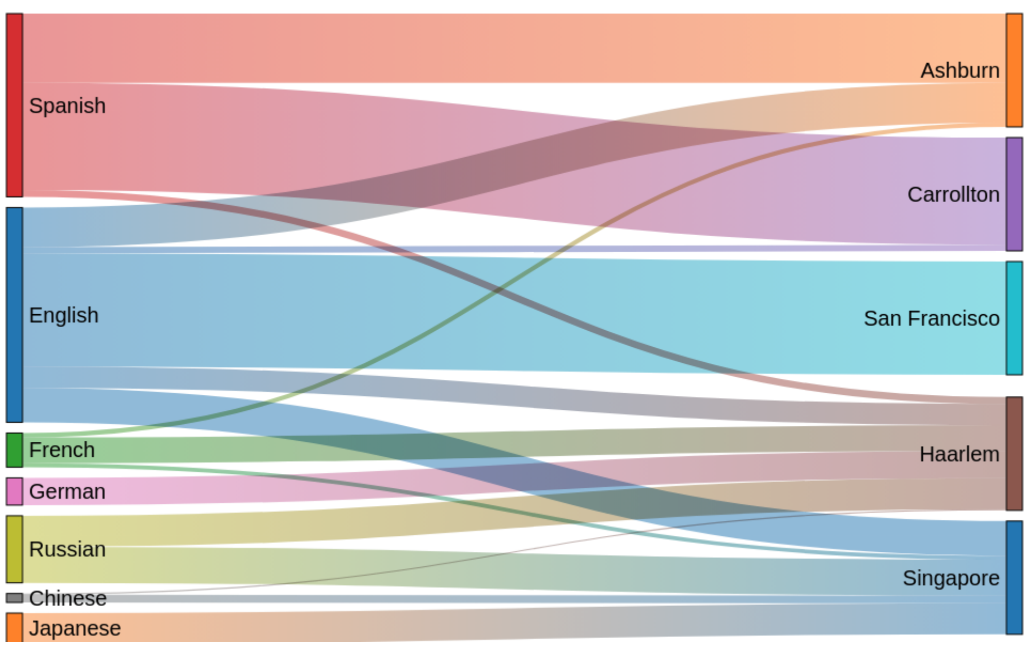

Figure 4.1: Language to data center flows. The thickness of the links at each data center site indicate the relative portion of traffic that the respective language demands at the data center.

4.3.2 System Boundary

The presented framework is intended to be used as a tool in data center life cycle assessments (LCA), where prevalent LCA practices are followed. In standard practice, product specific LCA studies generally entail four phases: 1) goal and scope phase, 2) the inventory analysis phase, 3) the impact assessment phase, and 4) the interpretation phase [15]. The methodology, data sets, and software tools presented in the research is conducive for the first three LCA phases which all require quantitative assessment of data center environmental footprints. Stated more explicitly, this research's goal is to provide a model that can be used to quantify the energy and carbon footprint of data center systems encompassing the embodied and operational phases in a single workflow.

Figure 4.2 illustrates the boundary conditions the boundary conditions with in the scope of this framework. In the scope are for paths of input of raw materials and energy from the ecosphere. Two paths lead to the embodied systems found in data centers; building construction and information technology manufacturing. While the other two path lead to the operational systems that are required to run the data centers.

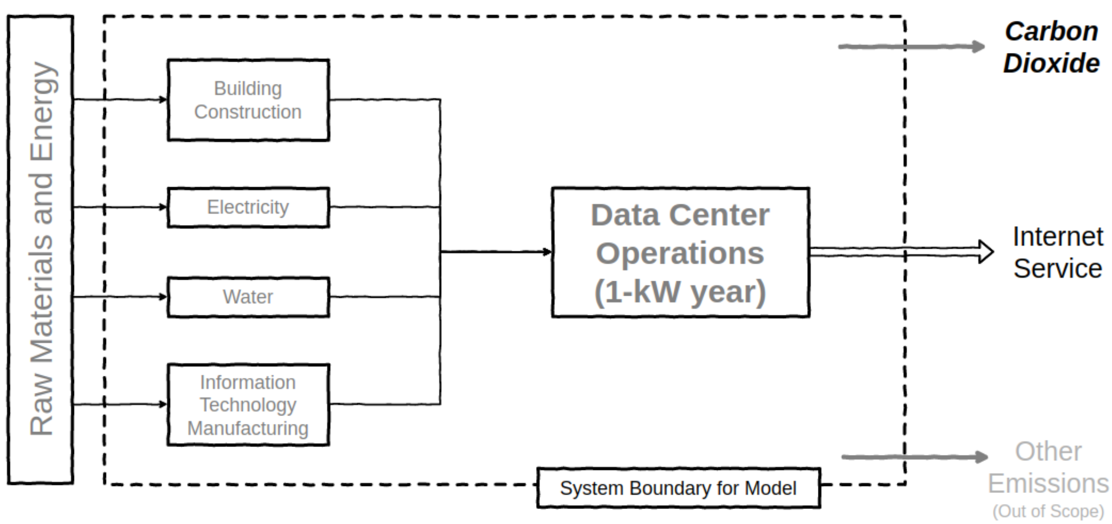

Figure 4.2: System Boundary for 1-kW of Provisioned Data Center Capacity.

The geographical bounds of the service are illustrated in Figure 4.3. The map indicates the countries in which each language from the Wikipedia set are the official languages. From Figure 4.1 it can be seen that one or more languages can be supported at a single data center facility. Although the workloads originate across international boundaries, the environmental emissions studied in this research are attributed to the data center sites only.

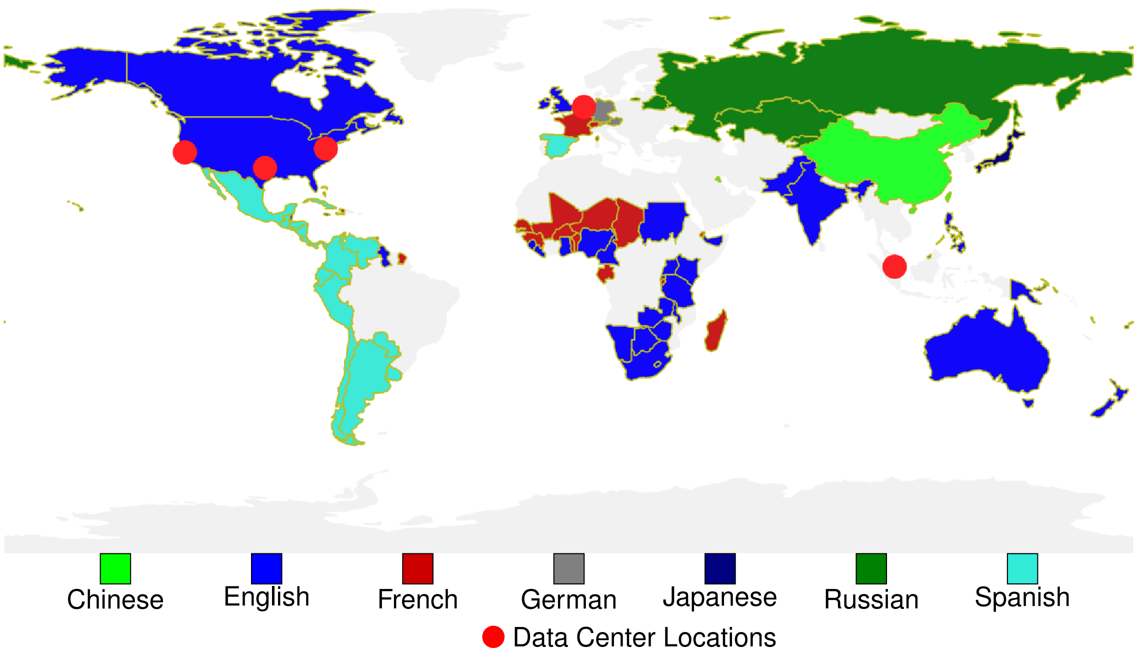

Figure 4.3: Source country of language and DC locations map.

Within these data center sites, specific sub-systems are segregated as listed in Table 4.1. Demarcation by these sub-systems allows development of the embodied and operational models to align with industrial cost breakdown data and in terms of EIO industrial vectors. Table 4.1 serves the structural backbone for all the models developed in this framework as discussed next.

System Boundary	**Examples**
Building Systems	Data Center Shell and Core
Cooling Systems	Air Handling Units and hydronic plants
Power Distribution	UPS, PDU, Cables, Diesel Generators
Information Technology	Compute, Storage, Network

Table 4.1: Sub System Boundaries

4.3.3 Operational Inventories

In this section, this research's methodology to assess the operational energy and carbon footprint of data center site infrastructure is presented. There are extensive theoretical data center energy use models in literature [16, 17]. These models can generally be segregated between IT and building systems domains. The model developed in this research couples the two domains by first selecting the right IT component to model; i.e. the IT component that the total energy of the data centers is most sensitive to.

Detailed industrial power usage insights are lacking in the public domain [18]. As an exception some industrial insights are provided by Barroso in [1, 19]. Barroso provides Google's distribution of power for various points-of-use as indicted by the pie-charts shown in Figure 4.4 for two generations of technology. In 2012, more than 80% of the energy in a data center was used by four components; CPU 42%, Cooling 15.4%, disk 14.3% and DRAM 11.7%. By 2017, the cooling overhead had decreased to only account for 3% a data center energy use, this decrease in cooling inflated the relative fraction of CPU and DRAM energy requirement. From these distributions, it is apparent that CPU's power usage is the dominant hot-spot.

The values in Figure 4.4 are annualized distributions. In practice the power demand of data centers is very dynamic and sensitive to complex workload dependencies. These dependencies lend themselves to be exploited in power proportional computing paradigms. Several proportional workload techniques are discussed in detail by O'Sullivan in [20]. This research takes such techniques and extends building energy models to be aware of proportional CPU loads based on coming network traffic to a data center site in an EnergyPlus model.

EnergyPlus provides a comprehensive indication of the operational energy and together with the marginal cost of energy model from Chapter 3, it exposes the carbon footprint for each of the five data centers and the language abstracted service pairs. Both models are based on the network coefficients from Chapter 1. The traffic coefficients are

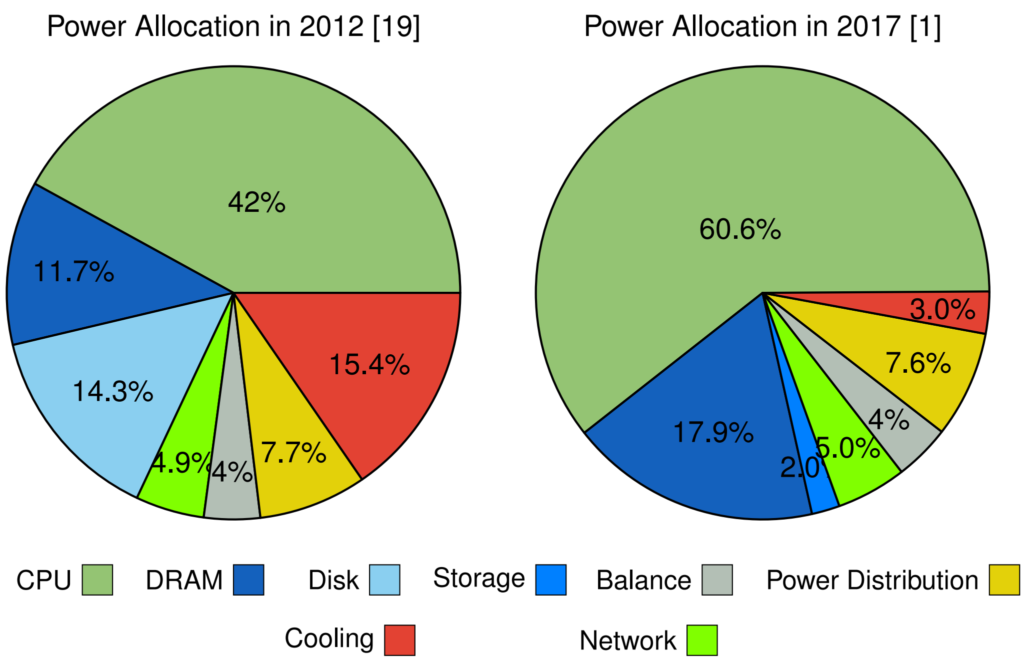

Figure 4.4: Data Center Power Distribution for Two Generation of Technology

used to reset the IT workloads at each data centers and the marginal costs to produce energy in the regional grid. The details of the model is indicated in Algorithm 1 and the general framework is illustrated in Figure 4.5.

The configuration of the model is as presented by Kumar in [7]; where the provisioned IT capacity of each data center is 492-MW. This provisioned capacity is scaled from the EnergyPlus simulation to represent metropolitan sized data center nodes. This scale sufficiently represents IT capacity provisioned at modern metropolitan area data center nodes. The specific data center wise allocation of each language is indicated in Figure 4.2.

4.3.4 Embodied Inventories

Embodied cost assessments account for the natural resources consumed when extracting, transforming, manufacturing, transporting, constructing, maintaining, and disposing the product under study. The first step in these assessments is to set the boundary

Algorithm 1 MEC coupled BEM algorithm

Require: $DOSCOE[region_i]$ & $traffic.language[site_{DC}]$
  for site and region in $DC.site$ and $DC.region$ **do**
    if traffic.language[site].all $! = 0$: **then**
      $D_{DC} \leftarrow BEM(site, traffic.language[site_{DC}])$
      $DOSCOE[region].demand += D_{DC}$
    end if
  end for
  return CO_2 footprint $=$ $GridSim(region, rps)$

Where:
$DOSCOE[region_i]$ = Dispatch Optimized System Cost of Energy Model for $region_i$ [21].
$site_{DC}$ = data center site.
$traffic.language_l[site_{DC}]$ = the traffic routed to data center for lanuage l.
$DC.site$ = list of all data centers in the network.
$DC.region$ = list of corresponding power grid region for the data centers in $DC.site$.
$DOSCOE[region].demand$ = demand vector of loads added to the power grid for every hour.

External Models:

$BEM(site, traffic.language[site_{DC}])$ is a EnergyPlus model of the data center. As an external argument the traffic profile for the language to the data center is passed. The traffic profiles serve as coefficients for the IT load, bounded by 0 and 1. The output is a vector indicating the building energy demands for each hour of the year.

$GridSim(region, rps)$ is a DOSCOE model with $DOSCOE[region].demand$ and renewable portfolio standard (rps) value indicating the required penetration percentage of renewable energy in the power supply. The model quantifies the costs of energy in terms of carbon footprint and monetary values.

conditions of the system. The second step is to select the method of the assessment. Two methods are available for such assessment; namely process based or EIO based. Process based methods are bottoms up procedures; accounting for all processes that are required to transform raw materials into consumer goods. However in process based methods as the product boundary and the count of processes to quantify can have exponentially correlations. This exponential growth can become an intractable task for building designers. A more tractable method is the EIO, where economic relationships between two industries at the national level are exploited.

Leontief's original use case of the EIO was to identify the contribution from all economic sectors required to produce a unit of specific products in an economy. The formulation of Leontief's EIO is indicated by Equation 4.1; where Y is the input costs, I is

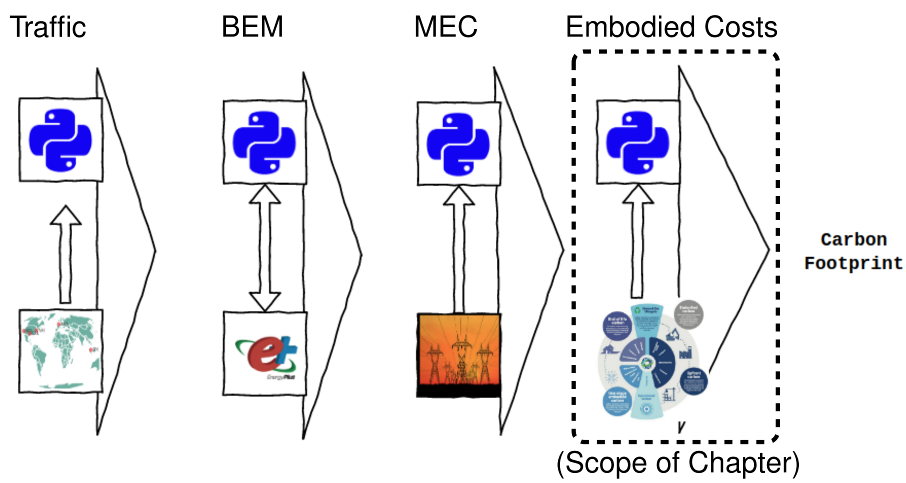

Figure 4.5: Model Process Flow Diagram.

Language	Site	Allocation %	Allocation MW
English	Ashburn	35	172
French	Ashburn	4	19
Spanish	Ashburn	61	301
English	Carrollton	5	27
Spanish	Carrollton	95	465
English	Haarlem	19	92
German	Haarlem	24	118
French	Haarlem	23	112
Chinese	Haarlem	1	5
Russian	Haarlem	27	134
Spanish	Haarlem	6	30
English	San Francisco	100	492
English	Singapore	30	150
Japanese	Singapore	27	135
French	Singapore	4	17
Chinese	Singapore	7	33
Russian	Singapore	32	157

Table 4.2: Provisioned Allocation of Language at each Data Center

the identity matrix, and A is the direct costs array [22]. In general direct costs represent the entire economy as a two-dimensional square array, where the rows and columns indicate distinct sectors within the economy. The values in the array indicate the row-index-sector's input into the column-header-sector. Since their development in the 1930's, EIOs have been commonly used to characterize national economic accounts. More recently, Matthews and Hendrickson showed that EIO can be used for environmental life cycle assessments as-well [22].

$$[I - A]^{-1}Y \tag{4.1}$$

Due to the streamlined approach of the EIO method compared to process based environmental assessments methods, the former is preferred for this research. Specifically, the United States Environmentally Extended Input-Out (USEEIO) model developed by Yang for the EPA is used in this research [23]. The USEEIO's economic relationship between sectors is based on 2007 cost data, while the environmental data are derived from 2013 emissions data. Both were the most up-to-date data available to the researchers at their time of publication. They suggest using 2013 cost of money for inputs, having demonstrated that the underlying structure of the economy has not shifted significantly between 2007 and 2013.

Building Systems

Typically, data center shells and structural cores are under the jurisdiction of the same building codes as other commercial buildings, i.e. Type IIA construction types. However, the dense population of computers along with the power and cooling requirements of data centers make several things clearly distinguishable from most commercial building types. The differences are significant enough to make the generic building sector's USEEIO environmental inventories invalid for data centers.

To more accurately represent data center buildings, several general contractor proposals for new data center construction and retrofit projects were reviewed for the cost distribution of the trades labor and equipment costs. The contractor cost allocations per trade divisions were then compared with the values of the handful of building construction sectors available in the USEEIO. Based on the comparisons, the USEEIO manufacturing buildings sector is found to be the most similar for data centers. However, there are still shortfalls between data centers and the manufacturing buildings sector. To overcome the shortfalls, a hybrid method was used to construct a more representative Y vector for data centers.

The resulting Y vector to scale for the building and building utilities systems is shown in Table 4.3. The table indicates the cost input required to provision a functional unit of 1 kW of data center capacity. Each sector's cost is derived from the contractor costs for the corresponding construction specification division as listed in their proposals. Seven explicit sectors are quantified in addition to the manufacturing building sector. The manufacturing building/us is a catch all sector. This sector catches all the building trades that are not represented by the other entries in the Y vector; such as concrete or steel.

ID	Sector Name	$/kW
333415	air conditioning, refrigeration, and heating equipment/us	550
333611	turbines and turbine generator sets/us	770
335311	specialty transformers/us	50
335313	switchgear and switchboards/us	100
335920	communication and energy wire and cable/us	100
335999	other miscellaneous electrical equipment and components/us	40
335314	relay and industrial controls/us	150
233230	manufacturing buildings/us	5250

Table 4.3: EEIO Costs Vector for Building and Utilities Infrastructure

Information Technology Systems

In this subsection the IT equipment's inventories are assessed. The IT equipment in a data center can be categorized broadly into three segments: compute, storage, and network. Within each of these segments, data center operators can further tune the hardware configurations to optimize their workloads. Figure 4.6 indicates component wise cost distribution for five hardware stock-keeping units (SKU). The SKU data were obtained under non-disclosure agreements from a large internet website operator.

The configuration of the server SKUs are optimized for the workloads that they support. For example the High-IO server is optimized for high rates of input and outputs;

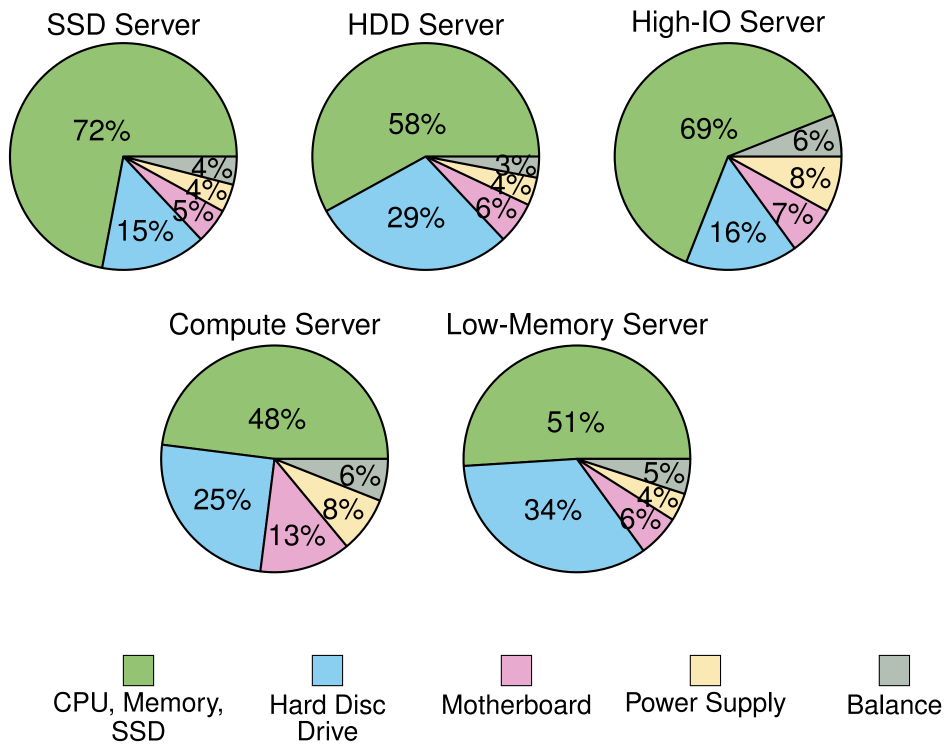

Figure 4.6: Cost apportionment of deployable hardware. See Table 4.5 for details of USEEIO sector mapping. The values are representative of deployment cost share per server during 2015 and 2016

where it processes information at high feed rates. As an example, this useful for workloads that require external communications with a large data set that is housed in a Mix-SSD server that has an abundance of data storage capacity. For a data center with dominant heterogeneous workloads will have a mix of of these SKUs, whereas a data center with a single workload would have only one or two distinct SKUs.

Server components such as processors (CPU), memory (MEM), and solid-state-drives (SSD) are fabricated using semiconductor manufacturing processes. These components are grouped together and are categorized as a single input into the USEEIO's semiconductor sector. Hard disc drives (HDD), motherboard (MBD), and power supply units (PSU) have representative sectors that allow a one to one mapping within the USEEIO. Other components such as chassis, thermal-heat sinks, and cable connectors are grouped into the balance category. The balance category is mapped to the generic computer sector in USEEIO. This costs apportionment values can be used to construct the Y vector for input in the USEEIO. However, before doing so the quantities of the components must be scaled up to by the count of servers provisioned in the data center.

Algorithm 2 is used to quantify the count of servers in each data center. Table 4.4 indicates the algorithm's power density value for each component found in each server. For each server the total power is calculated by summing the power for all components in it. The component apportionment for each server SKU is indicated in Table 4.5.

Algorithm 2 IT equipment Y vector algorithm

Require: P_{DC}, P_i, and C_i

$P_S \leftarrow \sum_{i=1}^{n} P_i$

$C_{ip} \leftarrow C_i \times f_i$

$S_{\#} \leftarrow \frac{P_{DC}}{P_S}$

$SC_{DC} \leftarrow S_{\#} \times P_S$

return $Y \leftarrow SC_{DC} \times C_{ip}$

Where:

P_{DC} = Provisioned Power of Data Center, (input variable)
P_i = Power demand for component i of server.
n = Number of components in the server.
C_{ip} = Provisioning cost of component i
C_i = Cost Apportions for server component i. See Figure 4.6.
f_i = Failure rate of component i. See Table 4.6.
$S_{\#}$ = Count of servers in data center
SC_{DC} = Total Server Costs for DC
Y = Vector of sector wise costs for input to USEEIO

Component	**Power (Watts)**	**Reference**
3.5" HDD	3.87	[24]
2.5" HDD	0.89	[24]
SDD	0.37	[24]
Single Socket CPU	105	[25]
Dual Socket CPU	210	[25]
Power Supply Unit	25	[17]
Motherboards	40	[26]
RAM	0.375	[27]
Fans (5 per server)	10	[26]

Table 4.4: Server component power distribution

Using the above characteristics of the data center systems, the embodied costs of the systems are amortized over their useful lives. Namely, the building systems are

Component	USEEIO Sector	SSD	Mix-HDD	High-IO	Compute	Low-Mem
CPU, Memory, SSD	semiconductors/us	72%	58%	69%	48%	51%
HDD	computer storage device readers/us	15%	29%	16%	25%	34%
Motherboard	printed circuit and electronic assembly/us	5%	6%	7%	13%	6%
Power Supply Unit	Electronic Capacitors and other components (except semiconductors and printed circuit assemblies)/us	4%	4%	4%	8%	4%
Balance of Parts	computers/us	3%	3%	3%	6%	4%

Table 4.5: Cost apportionment of deployable hardware stock keeping units (SKU)

Component	Failure Rate
CPU	0.03
MEM	0.27
HDD	0.67
SSD	2.56
MBD	1.57
PSU	0.30
BAL	0.25

Table 4.6: Information Technology Failure Rates [1, 2, 3].

amortized over 20 years, while the IT equipment is amortized over 3 years. For both of these systems the annual amortized costs are distributed on a hourly basis over the year. The hour distribution allows for the network traffic coefficients of each language and data center pair to apportion the embodied costs with the traffic coincident traffic for the hour.

As a final and important note about how the carbon footprint is translated from the USEEIO to a meaningful measure of sustainability. The USEEIO provides environmental impacts associated with the embodied costs of data center components based on the US Environmental Protection Agency's (EPA) Tool for Reduction and Assessment of Chemical and Other Environmental Impacts (TRACI) [28]. The impacts of the carbon footprint manifests itself as the global warming potential of the system. Therefore, the carbon footprint indicator from USEEIO is in terms of equivalent carbon dioxide mass.

4.4 Results

In this section the results of the methodology described above is provided. Again, the discussion is presented first by discussing the operational and embodied carbon footprints separately. Then those footprints are assessed in terms of the functional unit. Ultimately, the results of this research culminate by quantifying the carbon footprint of the data center based reference flow of provisioning 1 kW-year year of data center capacity.

4.4.1 Operational Carbon Footprint

Figure 4.7 shows CPU workload values for a single language at each of the data centers in the global network. The variance in the curve demonstrates the effectiveness of using the traffic coefficients. The CPU workloads are used in EnergyPlus as building plug loads that must be powered and cooled data center infrastructure. The EnergyPlus model performance a physics based simulation for the total building considering these CPU loads, the ambient environment, the interior operational temperatures, and systems inefficiencies to quantify the energy use for each hour of the year.

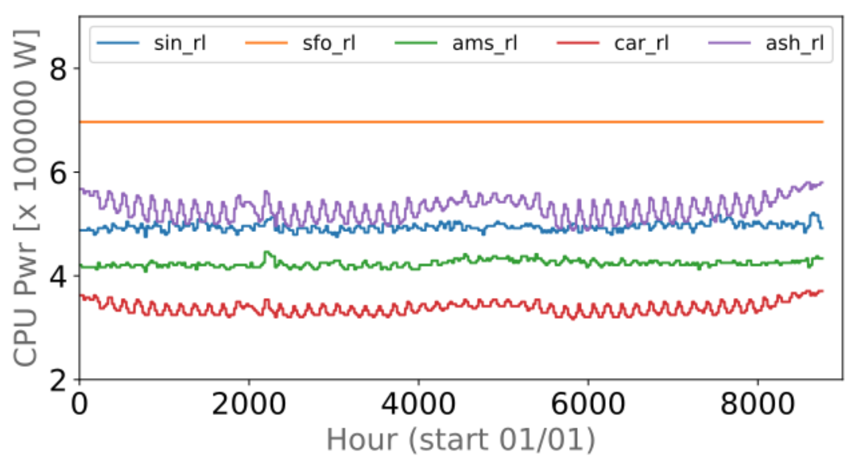

Figure 4.7: The curves indicates the CPU power demand by English language pages visits to each respective data center. The names of the curves are abbreviated for the reset load (rl) models run by Kumar in [6]. sin:Singapore, sfo:San Francisco, CA, USA, ams:Haarlem, Netherlands, car: Carrollton, TX, USA, ash: Ashburn, Virginia, USA

The energy demand of the building at an hourly interval from EnergyPlus is then passed to the DOSCOE - marginal costs of energy model. The model quantifies the data center's incremental cost of energy within it's power grid. The annualized carbon footprint of the English language pages is noted in Table **??**. c

Site	German	English	Spanish	French	Japanese	Russian	Chinese	Total
California	0	1 301 549	0	0	0	0	0	1 301 549
Ercot	0	519 466	1 341 227	0	0	0	0	1 860 693
Midatlantic	0	563 085	1 109 162	661 919	0	0	0	2 334 165
Netherlands	464 780	470 714	473 987	766 542	0	1 027 424	457 393	3 660 841
Singapore	0	591 300	0	614 150	873 602	860 595	578 499	3 518 145
Total	464 780	3 446 113	2 924 376	2 042 611	873 602	1 888 019	1 035 892	12 675 393

Table 4.7: Data Center Operation Carbon Footprint

4.4.2 Embodied Carbon Footprint

The second part of the modeling framework quantified the embodied cost of each data center using the USEEIO. In Figure 4.8, the operational carbon footprint from the above BEM models are compared with the embodied carbon costs through USEEIO framework. In the figure the left curve shows the operational carbon footprint over the yearly model, while on the right the embodied costs of all SSD servers (see Figure 4.6) of the same language - data center pairs are shown. Note the scale of the two graphs which clearly indicate that the embodied costs dominate over the operational costs. The total costs with both the operational and embodied models are shown in Figure 4.9.

4.4.3 Cost of Process flow per per functional unit

In this research, the focal point for data center sustainability is its total carbon footprint. In the preceding methodology, the carbon footprint over a year has been quantified with consideration of the operational and embodied life cycle phases of the data center. However, those footprints have not yet been correlated with the functional unit of performance for the data center system. To correlate with the functional unit, the provisioned capacity of the network of data centers to support each language is needed.

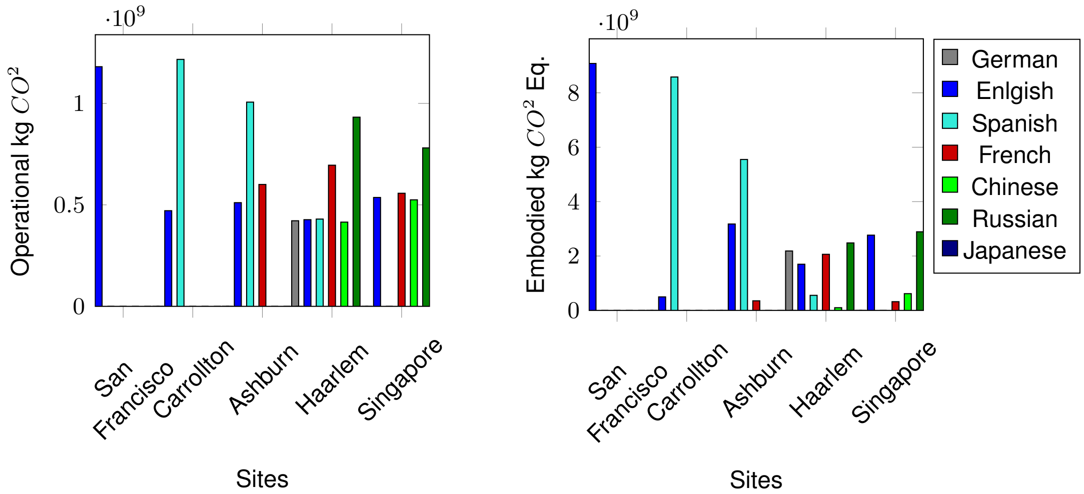

Figure 4.8: Carbon footprint of pages in each language at each data center. Left - Operational Energy for a year. Right - Embodied energy for a year from TRACI Impact Potential Global Climate Change.

The process flow of CO^2 in kg per year of provisioned kW capacity for each language is shown in Figure 4.10. The values are sorted in ascending order form left to right; where Spanish has the lowest process flow values and the Chinese has the highest values. The lower process flow values are for languages that have balanced distributions to the global data center fleet, whereas the high process flow values are for languages which the Singapore data center has a relatively significant share of its traffic (see Figure 4.1).

However, the most significant attribute that appears to indicate the carbon footprint per functional unit is the total data center power footprint. This is demonstrated by the small power footprint of the Chinese language incurring the highest carbon footprint costs. These results reflect all the data centers with a single server SKUs, namely the SSD Servers. Nonetheless, the model can be extended to other servers types or a combination of server types.

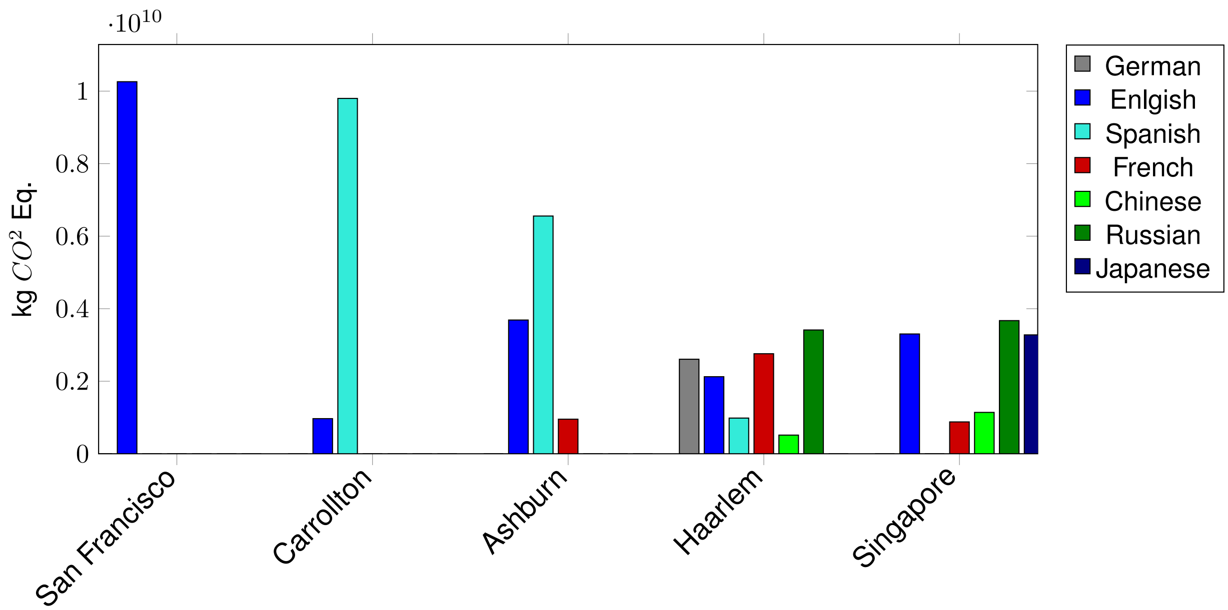

Figure 4.9: Total Carbon Footprint of each language at each Data Center

4.5 Conclusion

This research coupled process and economic input-output based methods using a simple model that quantifies the operational and embodied carbon footprint of a global network of data centers. The presented model provides two key insights. First, he model shows that generally the embodied carbon dioxide of data center materials greatly exceeds the carbon dioxide emissions during the operations. Second, the model also shows that the operational carbon footprint varies based on data center locations; specifically the local ambient weather and the energy mix of the local power utility grid. These two insights indicate that while the focus on operational efficiency can lead to reductions in global warming potential from carbon dioxide emissions, there is still a long way to go to make data centers truly sustainable over its entire life-cycle. This research provides a starting point which data center architects and operators can use to assess different combinations of data center deployment strategies; addressing geographical location, network load balancing, IT hardware configurations, and building designs.

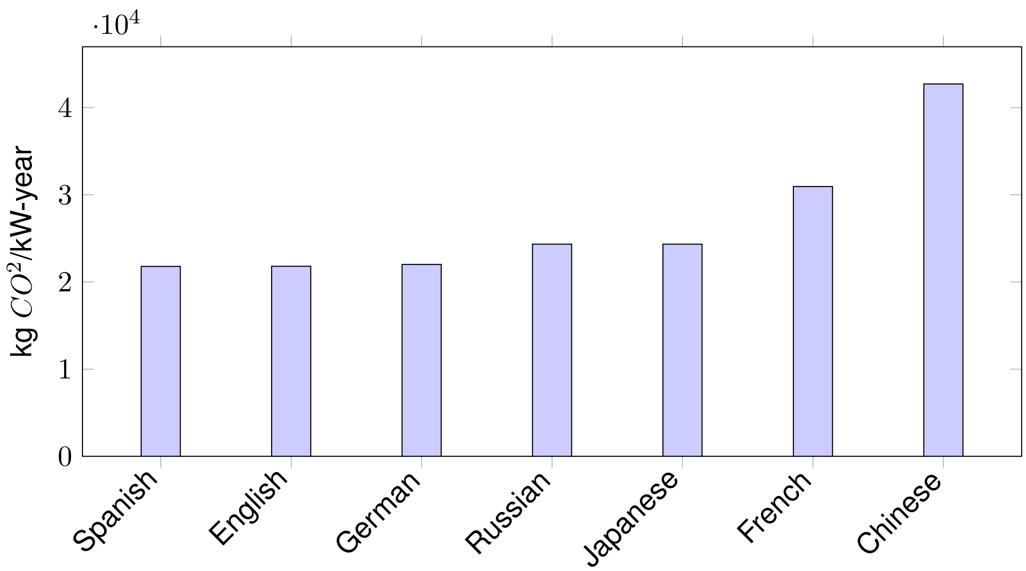

Figure 4.10: Carbon Footprint per Functional Unit of Process Flow

Bibliography

[1] L. A. Barroso, U. Hozle, and P. Ranganathan, *The Datacenter as a Computer: Designing Warehouse-Scale Machines 3rd ed*. Morgan and Claypool, 2018.

[2] P. Kogge, *ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems*. DARPA, Washington, DC, 2008.

[3] S. A. J. Srinivasa, "The case for lifetime reliability-aware microprocessors," in *The 31st International Symposium on Computer Architecture (ISCA-04)*, 2004.

[4] A. Shehabi, "United States Data Center Energy Usage, United States Department of Energy," 2016. [Online]. Available: https://eta-publications.lbl.gov/sites/default/files/lbnl-1005775_v2.pdf

[5] A. Shah, C. Bash, R. Sharma, T. Christian, and C. Patel, "Evaluating life-cycle environmental impact of data centers," vol. 133, pp. 031 005–1, 2011.

[6] E. Kumar, "Towards energy simulations for proportionally designed and controlled data centers," in *Accepted for Building Simulation and Optimization 2020*, 2020, loughborough (UK), 21-23 September 2020.

[7] ——, "Wide area network based data center energy simulations for internet services," in *Accepted for 11th International Conference on Improving Energy Efficiency in Commercial Buildings and Smart Communities*, 2020, frankfurt (Germany), October, 2020.

[8] B. Whitehead, "The life cycle assessment of a UK data centre," in *International Journal of Life Cycle Assessments*, vol. 20, 2015, pp. 332–349.

[9] S. Boyd, *Life-Cycle Assessment of Semiconductors; Dissertation Submittal*. University of California Berkeley, 2009.

[10] M. Alcaraz, A. Noshadraven, M. Zgola, and R. Kirchain, *Streamlined life cycle assessment: A case study on tablets and integrated circuits*. Journal of Cleaner Production, no, 2018.

[11] C. F. Murphy, G. A. Kenig, D. T. Allen, J.-P. Laurent, D. E. D. E. Science, and Technology, *37 (23)*. 5373-5382 DOI: 10, 2003.

[12] D. Kline, A. Jones, and N. Parshook, "Holistically evaluating the environmental impacts in modern computing systems," in *2016 Seventh International Green and Sustainable Computing Conference (IGSC), Hangzhou, China*, 2016.

[13] LBNL, *Lawrence Berkeley National Lab and North Western University Cloud Energy and Emissions Research Model, CLEER. 2013, Retrieved: May 05, 2020*, 2016, https://www.cleermodel.lbl.gov/.

[14] T. The Green Grid, "Data center life cycle assessment guidelines. Retrieved: May 13, 2020," 2012, https://www.thegreengrid.org/en/resources/library-and-tools/236-Data-Center-Life-Cycle-Assessment-Guidelines.

[15] I. S. Organization., *Environmental management - Life Cycle Assessment - Principles and Framework*. Switzerland: International Standards Organization, 2006.

[16] M. Dayarathna and Y. Wen, "Data center energy consumption modeling: A survey," *IEEE Communications & Tutorial*, vol. 18, no. 1, pp. 732–794, 2016.

[17] Y. Joshi and P. Kumar, *Energy Efficient Thermal Management of Data Centers, New York: Springer Science and Business Media*, 2012.

[18] E. Masanet, A. Shehabi, N. Lei, S. Smith, and J. Koomey, "Recalibrating global data center energy-use estimates," *Science*, vol. 367, no. 6481, pp. 984–986, 2020. [Online]. Available: https://science.sciencemag.org/content/367/6481/984

[19] L. Barroso, U. Hozle, and J. Clidaras, *The Datacenter as a Computer: Designing Warehouse-Scale Machines 2nd ed*. Morgan and Claypool, 2013.

[20] N. O'Sullivan, U. Ammu, and M. Yazdani, "Energy proportional power management for data center applications," *DesignCon*, 2015.

[21] J. Platt, O. Pritchard, and D. Bryant, "Data center construction market 2019-2024 | global industry overview by size, share, future demand, latest research by competitors, segmentation and regional forecast," *SSRN*, vol. 08, 2017.

[22] S. Matthews, C. T. Hendrickson, and D. H. Matthews, "Chapter 8: LCA Screening via Economic Input-Output Models," in *Life Cycle Assessment: Quantitative Approaches for Decisions That Matter – lcatextbook.com*, 2015.

[23] Y. Yang, W. Ingerwersen, T. Hawkins, M. Sroka, and D. Meyer, "USEEIO: A new and transparent United States environmentallyextended input-output model," *Journal of Cleaner Production*, vol. 158, pp. 308–318, 2017.

[24] H. Fuchs, A. Shehabi, M. Ganeshalingam, and L. B., "Descroches," *B*, vol. servers, 2019.

[25] Kaggle, *Computer Parts Dataset. Retrieved: April 06, 2020*, 2017.

[26] www.buildcomputers.net, *Power Consumption of PC Components in Watts, Retrieved: May 05, 2020*, 2020, http://www.buildcomputers.net/power-consumption-of-pc-components.html.

[27] www.crucial.com, *How much power does memory use, Retrieved: April 04, 2020*, 2020, https://www.crucial.com/support/articles-faq-memory/how-much-power-does-memory-use.

[28] U. S. E. P. Agency, *Tool for Reduction and Assessment of Chemicals and other Environmental Impacts, TRACI. Retrieved: April 05, 2020*.