Politecnico di Torino

Master's Degree in Electrical Engineering

March 2025

Exploiting Free cooling in Data Centre Application

Politecnico di Torino

Author:
Federico Mascitelli

Supervisor:
Prof. Paolo Lazzeromi

Abstract

The era of digitization has taken full throttle, and we have become increasingly dependent on the virtual world and the information services it offers, spanning from cloud computing to artificial intelligence and large-scale data analytics.

This growing demand has led to a corresponding increase in the energy consumption of data centers, which serve as the backbone of modern information structures. Additionally, the push for net-zero emissions scenarios has put pressure on various sectors, including data centers, to optimize their energy use and integrate greener sources into their energy mix.

Proper server temperature regulation is critical for data centers, as overheating can significantly impact their quality of service offered. Furthermore, the load of electrical energy required for cooling of data centers eats up a big piece of the total slice of overall load demanded.

This thesis explores the potential advantages of combining free cooling techniques and load management strategies to a hypothetical data center located in the colder regions of Europe, specifically Ireland. The free cooling technique is applied to the computer room air conditioning systems. The load management strategy is implemented by taking advantage of low-usage periods and the capability to distribute tasks across geographically dispersed data centres experiencing contrasting climates — shifting load away from warmer locations where possible — all while ensuring the quality of service promised to users within a specified proximity to each designated data centre.

Through the application of an optimization tool, XEMS13, this study aims to evaluate the results of the analysis and investigate the viability of a real life application considering the economic, energetic, and environmental benefits of the proposed cooling and load management approach.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Increasing pressure is being placed on reducing energy consumption across all sectors due to the escalating climate crisis. This urgency drives a widespread push for optimization, where every process, system, and facility is scrutinized to extract maximum efficiency from available resources. Data centers, as significant energy consumers, are no exception, with operators being compelled to adopt innovative technologies, smarter load management strategies, and improved cooling techniques to minimize their environmental footprint while maintaining performance and reliability.

The demand for digital services has been growing at an extraordinary rate, driven by the rapid expansion of cloud computing, video streaming, e-commerce, artificial intelligence applications, and the proliferation of connected devices. Since 2010, the global population of internet users has more than doubled, while overall internet traffic has skyrocketed by a factor of 25. Despite this exponential growth, substantial advances in energy efficiency — in both data centres and data transmission networks — have helped curb the increase in overall energy consumption. Thanks to innovations in server hardware, cooling systems, network optimization, and improved software management, the combined electricity demand of these two sectors currently accounts for approximately 1% to 1.5% of global electricity consumption. This relatively modest share, considering the scale of digital transformation, highlights the crucial role that technological innovation plays in ensuring sustainable digital infrastructure.[8]

Digital technologies influence global energy consumption and carbon emissions both directly and indirectly. On the direct side, data centres, networks, and connected devices all consume electricity, meaning their environmental impact is closely tied to the carbon intensity of the electricity grids they rely on. Data centres located in regions where renewables and low-carbon energy sources make up a greater share of generation tend to have significantly lower associated emissions, while those dependent on fossil fuel-heavy grids contribute more substantially to global greenhouse gas emissions.

Beyond their direct footprint, digital technologies play a pivotal role in shaping and accelerating the global energy transition. Through advanced data analytics, smart grids, real-time monitoring, predictive maintenance, and demand-side management systems, digitalisation enhances the efficiency, flexibility, and resilience of modern energy systems. These tools enable the integration of variable renewable energy sources, improve energy forecasting, and optimise energy consumption across industries and sectors.

However, this digital transformation is a double-edged sword. While it can support decarbonisation and boost efficiency, the rapid proliferation of data-intensive services (such as streaming, cloud computing, blockchain operations, and AI training models) could drive up demand for electricity, offsetting some of the gains achieved through improved efficiency. Therefore, aligning digital innovation with decarbonisation strategies is essential to ensure that digitalisation becomes an enabler — rather than a barrier — to achieving climate goals.

The overarching objective, as in many other sectors, is to achieve net zero emissions by 2050. This goal aligns with international climate commitments aimed at limiting global warming and reducing reliance on fossil fuels.

The growing energy intensity of the data centre sector is reflected in global electricity consumption estimates for 2022, which ranged between 240 and 340 TWh — equivalent to approximately 1% to 1.3% of total global electricity demand, excluding the energy required for cryptocurrency mining. Data centres and data transmission networks collectively contribute approximately 1% of global greenhouse gas emissions linked to energy use.

While substantial efficiency improvements have been made, these gains have been outpaced by the rapid expansion of large-scale data centres. This segment alone has seen energy consumption rise by roughly 20% to 40% per year, driven by the ever-increasing demand for cloud services, AI applications, and high-performance computing.

The combined electricity usage of major hyperscale operators — including Amazon, Microsoft, Google, and Meta — more than doubled between 2017 and 2021, reaching approximately 72 TWh in 2021. This trend underscores the sector's growing impact on global energy systems and the critical need for ongoing efficiency gains, sustainable energy sourcing, and smarter workload distribution to mitigate future consumption and emissions growth.

Several smaller countries that are experiencing rapid growth in their data centre sectors are also witnessing significant surges in electricity demand directly linked to these facilities.

reland serves as a prime example, where electricity consumption from data

centres has more than tripled since 2015, reaching approximately 18% of the country's total electricity demand in 2022. If this expansion trend continues — particularly alongside growth from other large non-industrial energy consumers — projections indicate that by 2031, this combined group could account for as much as 28% of Ireland's total electricity demand, unless substantial new generation capacity is added to the grid.[8]

A similar pattern is emerging in Denmark, where electricity consumption from data centres is projected to increase sixfold by 2030, ultimately representing nearly 15% of the nation's electricity demand. These cases underscore the growing regional pressures that the data centre industry places on national energy systems, highlighting the urgent need for both capacity planning and the integration of low-carbon energy sources to balance growth with sustainability goals.

First and foremost, it is essential to understand what data centres are. A data centre not only provides the infrastructure to store vast and ever-increasing amounts of data, but also ensures its protection and guarantees continuous data processing. To meet these requirements, a data centre can be viewed as a complex information infrastructure, composed of servers, storage systems (including filing systems), uninterruptible power supplies (UPS), routers, and various other components — all working together to ensure business continuity.

Thus, the Data Center is a specialized facility, or a network of interconnected facilities, dedicated to the centralized housing, interconnection, and operation of information technology (IT) and network telecommunication equipment. These facilities provide essential data storage, processing, and transport services, which form the backbone of modern digital infrastructure. To ensure reliable and continuous operation, the data centre must also include comprehensive power distribution systems, advanced environmental control mechanisms, and robust security systems, all carefully designed to deliver the desired levels of service availability, redundancy, and operational resilience.

1.1 Archeticture and Layout of a Data centre

A Data Processing Centre (DPC) is a facility designed to house and manage interconnected systems, with its primary objective being to ensure security, reliability, and operational continuity. These facilities are responsible for maintaining uninterrupted processing, safeguarding stored data, and preserving optimal environmental conditions to support the proper functioning of critical infrastructure.

The preservation of this balance, which combines security, resilience, and redundancy, is also carefully considered in the physical design of the data centre. A clear separation is maintained between the data halls, where IT equipment is housed, and the infrastructure that supplies power, cooling, and other essential services. This physical division helps reduce risk and ensures greater reliability. A typical layout, illustrating all the key components involved, can be seen in Figure 1.1

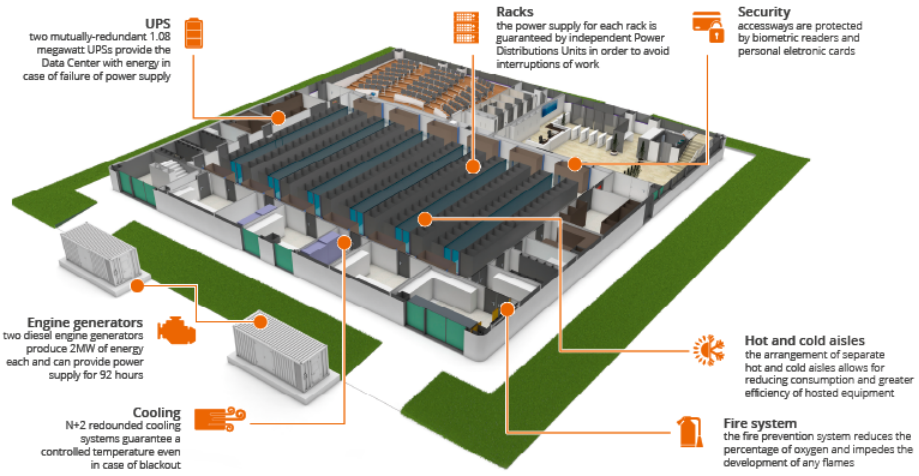


Figure 1.1: Layout of a Data Centre. Source: Vianova

Beyond the IT equipment itself, a fully functional data centre relies on an intricate ecosystem of supporting infrastructure, which can be broken down into several key domains:

- **Power Distribution Systems:** Reliable and redundant electrical infrastructure ensures that IT equipment receives stable power. This includes high-capacity transformers, switchgear, power distribution units (PDUs), and uninterruptible power supplies (UPS) that provide backup

power during grid failures. In some cases, onsite generators provide extended backup.

- **Cooling and Environmental Control Systems:** Servers generate a significant amount of heat, requiring advanced thermal management solutions to maintain optimal operating conditions. These systems may include precision air conditioning, chilled water systems, free cooling setups, and containment strategies that separate hot and cold air streams for efficiency. Maintaining acceptable humidity levels is also critical to prevent electrostatic discharge and equipment corrosion.

- **Network Infrastructure:** Modern data centres serve as key nodes in global and regional networks, requiring high-capacity fiber links, routers, switches, and interconnects to ensure seamless data flow between facilities, customers, and cloud services. Many data centres also function as internet exchange points (IXPs), where different networks interconnect directly to optimize traffic routes

- **Physical and Cyber Security Systems:** Ensuring physical access control is essential, often involving biometric access controls, surveillance systems, intrusion detection, and 24/7 onsite security personnel. In addition to physical security, cybersecurity infrastructure protects against data breaches, malware, and distributed denial-of-service (DDoS) attacks, ensuring data confidentiality, integrity, and availability.

- **Service Availability and Redundancy:** A defining characteristic of modern data centres is their focus on high availability, typically expressed through service level agreements (SLAs) with guaranteed uptime percentages (e.g., 99.999%). Achieving this requires redundant power, cooling, and networking, as well as robust disaster recovery and failover capabilities between geographically distributed facilities.

- **Scalability and Flexibility:** As demand for data storage and processing grows, data centres are designed with modular or scalable architectures that allow for the rapid addition of compute, storage, and networking resources. This flexibility allows operators to respond to both short-term spikes in demand and long-term growth trends.

- **Sustainability Considerations:** In recent years, environmental concerns have driven data centres to adopt more energy-efficient cooling systems, renewable energy sources, and heat recovery technologies. Metrics such as Power Usage Effectiveness (PUE) and Carbon Usage Effectiveness (CUE) are increasingly used to benchmark sustainability performance.

1.2 Data Centre Classification

Classifying a data centre involves the evaluation of several critical factors that collectively determine its reliability, availability, and overall operational performance. These classifications are often standardized to establish a common understanding of infrastructure resilience and the facility's capacity to sustain operations under varying conditions.

1.2.1 Tier Classification

The most commonly used classification system is the Tier Classification System, developed by the Uptime Institute. This system categorizes data centres into four tiers, each representing a different level of redundancy, fault tolerance, and availability. [9]

1. Tier I: Basic infrastructure with limited redundancy, offering approximately 99.671% availability. This type of data centre has a single path for power and cooling, making it vulnerable to failures and requiring planned downtime for maintenance.
2. Tier II: Redundant components improve reliability, resulting in around 99.741% availability. While there is some redundancy for power and cooling, the system still relies on a single distribution path, limiting flexibility.
3. Tier III: Concurrently maintainable infrastructure, providing 99.982% availability. These data centres have multiple paths for power and cooling, allowing equipment to be maintained or replaced without disrupting operations.
4. Tier IV: Fault-tolerant infrastructure, ensuring the highest level of availability at 99.995%. These data centres have fully redundant systems and paths, allowing any component to fail without affecting overall operations.

1.2.2 Classification by Size and Scale

In addition to the Uptime Institute's Tier Classification System, data centres can also be categorized according to their size and operational scale, which defines the breadth of their infrastructure, total processing capacity, and the nature of the organizations they support. These classifications offer valuable insight into the intended functional role of each data centre, influencing design parameters, resource allocation strategies, and the level of operational complexity required to meet performance and reliability expectations.

Enterprise Data Centres

Enterprise data centres are dedicated facilities owned and operated by a single organisation, serving to support its internal IT infrastructure and business functions. These data centres are usually designed to meet the organisation's specific requirements for data storage, application hosting, and internal networking. Their size, configuration, and technical setup are customised to align with the organisation's operational needs, whether it's a financial institution, a manufacturing firm, or a government body.

They are typically moderate in size, with power capacities ranging from 100 kW to 5 MW, depending on the organization's requirements.

Electrical design often prioritizes reliability and compliance, incorporating backup generators (optional), uninterruptible power supplies (UPS), and redundant power distribution paths, particularly for critical business functions.

Colocation Data Centres

Colocation data centres are shared facilities where multiple organisations lease space, power, cooling, and network services from a third-party provider. Within these facilities, tenants are responsible for installing and managing their own servers and networking hardware, while the colocation provider ensures physical security, environmental controls, and redundant power and network infrastructure.

The key benefit of colocation data centres lies in their cost-effectiveness, as businesses can take advantage of shared infrastructure, gaining access to high-performance data centre environments without the substantial upfront investment required to construct and maintain their own facilities. In addition, colocation providers generally offer flexible scalability options, allowing tenants to increase their allocated space and resources as their IT needs evolve.

Colocation facilities are widely used by small and medium-sized businesses as well as large multinational corporations needing regional infrastructure to support content delivery, data backups, or disaster recovery strategies.

These facilities range from medium-sized regional hubs to larger metro data centres, with power capacities typically between 1 MW and 50 MW.

Hyperscale Data Centres

Hyperscale data centres are massive facilities designed to support extremely large-scale computing environments, typically operated by cloud service providers, content delivery networks (CDNs), and large technology enterprises. These facilities are optimized to host vast numbers of servers, enabling the delivery of cloud computing services, big data analytics, artificial intelligence (AI) workloads, and global-scale applications.

Hyperscale data centres are characterized by: extensive automation, custom hardware and software stacks, highly optimized power and cooling systems to maximize energy efficiency and geographically distributed architectures to support global users with low-latency access.

They are engineered for extreme efficiency and high-density computing, with power capacities ranging from 50 MW to over 300 MW. Some of the world's largest hyperscale campuses exceed 1 GW of total power capacity when multiple buildings are combined.

1.3 Cooling and free cooling in data centres

For a data centre to provide uninterrupted service, in addition to having continuous power to supply the racks of servers, the conditions within the server rooms must be continuously monitored and met with specific conditions (e.g. temperature, humidity and air circulation within the server rooms). If these conditions are not met, the servers can fail or shut down and, in the worst case, receive irreversible damage. This is why this section within the data centre architecture is essential.

The cooling load attributable to IT equipment exceeds 90% of the total indoor air conditioning in all climate zones.[10] Two different cooling methods tackle this demand: direct liquid cooling and air cooling.

1.3.1 Direct liquid cooling

Direct liquid cooling uses several strategies, all having as a common denominator the characteristic of transferring waste heat from a point to a fluid at or near the said point rather than transferring it to room air [1].The strategies utilised are the following:

1. **Direct-to-chip cooling** – Integrates the cooling system directly into the computer's chassis. The refrigerated liquid is pumped through a network of cold plates placed in direct contact with components like CPUs, GPUs, or memory cards.
   As the fluid flows through the plates, it absorbs heat generated by heat-generating components. The now-heated liquid is transported to a cooling system or heat exchanger, which releases the absorbed heat and cools it down. Once cooled, the fluid is recirculated to the cold plates to repeat the process.

2. **Rear-door heat exchangers** – Heat exchangers are mounted at the back of the server rack instead of its back door. Server fans push the warm air through a heat exchanger, where the heat is effectively dissipated. The liquid circulates in a closed-loop system, continuously transferring heat away from the components and facilitating the cooling process.

3. **Immersion cooling** — The newer of the three strategies listed submerges all internal server components in a nonconductive dielectric fluid, sealed in a container to prevent leakage. This process requires less energy than the other approaches. The coolant is continuously circulated and cooled to dissipate the heat.

Liquid is a better heat conductor than air, which explains why this cooling method will be a key part of high-power-intensity AI data centres. The major con of liquid cooling is the potential damage it could cause to the servers, with the risk of leaking and devastating the hardware.

For the scope of this thesis, we will focus on air cooling, as it is the most utilised technique today and can be utilized much more easily with the free cooling technique.

1.3.2 Air cooling

Air cooling has been utilised since the inception of data centres. Cold air is directed across or around the hardware, replacing warmer air with cooler air to remove heat through an exchange process. The cold air can flow from a raised floor configuration, seeping through the perforated floor tiles or from classic vents on top of the equipment.

Air management

Air management components in data centres has as its primary target making sure that the exhaust air is not mixed with the cool supply air.

Effective air management reduces cooling air bypassing it around the rack intakes and prevents recirculation of hot exhaust air back into those intakes. This approach has the advantage of being able to increase the supply air temperature while maintaining optimal operating conditions for IT equipment.

If properly implemented, air management systems offer various benefits, including reduced operating costs, lower initial investment in cooling equipment, increased power density within the data centre, and a decreased risk of interruptions or failures caused by heat-related processing.

Major considerations to be aware of in air management design are: the configuration on the equipment of the air intake and exhaust ports, the placement of supply and return vents, understanding the large-scale airflow patterns within the space and the temperature set points to obtain a specific airflow [1].

A crucial aspect of air management is the arrangement of the racks. Conventional layouts utilise hot aisle/cold aisle configurations. As the name implies, it is nonetheless alternating rows of racks, with cold aisles designated as the side to intake the cool air and hot aisles serving as the heat exhaust side. This design is intended to optimize airflow and significantly enhance the air-side efficiency of the cooling system.

All equipment within the racks are configured for a front-to-back airflow pattern. Cool air is drawn in from the cold aisles in front of the racks, while hot air is expelled into the hot aisles located behind them.

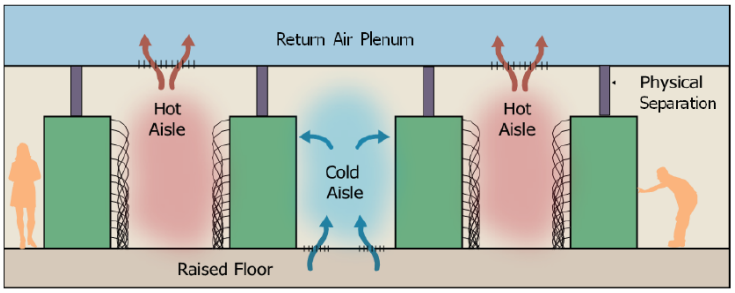


Figure 1.2: Hot/cold aisle configuration, with physical barriers. Supply air is provided from the raised floor and hot air is exhausted in the roof vents. Source: [1]

A further improvement in efficiency is obtained when the rows of racks are positioned back-to-back and any vacant slots on the intake side of the racks are sealed off. These barriers, acting as obstacles constraining two separate zones, a cold one and a hot one, help reduce the recirculation of hot air into the cold aisles ensuring more effective cooling and a stable thermal environment. With effective isolation, the temperature in the hot aisle no longer affects the racks or the reliable operation of the data centre, as the hot aisle functions solely as a heat exhaust.

The air-side cooling system is specifically designed to deliver cold air exclusively to the cold aisles while drawing return air solely from the hot aisles, thus ensuring efficient and well-regulated airflow management. Airflow management can even be pushed further from a room-based air cooling system to a row-based or rack-based one.

Cooling unit: CRAC and CRAH

There are two possible choices for cooling units: CRAC (Computer Room Air Conditioner) and CRAH (Computer Room Air Handler).

Suppose a traditional air conditioning system were to come to mind and the principles it works by it can be associated to a CRAC unit. CRAC units operate in the same way but are specifically designed to anage temperature, airflow, and humidity within data centre computer rooms. Making use of a direct expansion refigeration cycle, where air is cooled passing over a coil filled with a refrigerant. Once this refrigerant has expelled the absorbed heat, it stays cool, passing by a compression phase. The previously absorbed heat is expelled through the refrigerant, which can be glycol, water, or ambient air, depending on the system's configuration.

Like CRAC systems, CRAH units use fans to circulate air over cooling coils

to remove excess heat. However, instead of refrigerants, CRAH units rely on chilled water. This chilled water is supplied by a separate chiller or chilled water plant. Warm air from the computer room is drawn into the unit, passing over the chilled water coils, transferring heat from the air through the heat exchangers to the water. The warmed water running in the heat exchangers is then returned to the chiller to be cooled. With adjustable fan speeds, CRAH units can maintain stable temperature and humidity levels while offering greater operational flexibility.

The key difference between CRAC and CRAH units is their cooling mechanisms. CRAC units rely on refrigerants and compressors, whereas CRAH units use chilled water and control valves. CRAC units, once activated, typically run in a fixed mode, making the modification of cooling levels a task in case there was to be a change in demand in the computer room.[11]
While their operation is straightforward, CRAC systems have more components, leading to a greater need for regular maintenance and an increased risk of component failures. Proper upkeep is crucial for ensuring that CRAC units perform optimally. They are highly reliable and best suited for data centers with electrical loads of less than 200 kW and lower availability requirements.

In contrast, CRAH units are generally more energy-efficient. Their cooling cycle enhances heat removal while maintaining a similar energy footprint to CRAC units. Designed for data centers with electrical loads of 200 kW or more and moderate to high availability needs, CRAH units become even more efficient as data capacity increases, making them the preferred choice for larger data centers.
Both CRAH and CRAC have the possibility of integrating within their systems water-side and air-side economizers.

1.3.3 Free Cooling

Free cooling, also known as economizer cooling, is an energy-efficient technique that uses ambient outdoor air to cool the interior of a building, thereby reducing the reliance on traditional mechanical cooling systems.[12]

The concept of free cooling is based on the principle that in certain climates, the outdoor air temperature can be sufficiently low to directly cool the interior of a building without the need for energy-intensive refrigeration systems.

In the data center industry this technique is widely used. For instance, in Nordic countries with cold climates, free cooling is heavily utilized, allowing cool outdoor air to directly cool facilities for much of the year. Similarly, data centers in coastal regions with mild temperatures take advantage of the ocean breeze to reduce reliance on mechanical cooling systems.

Free cooling has been successfully adopted in various data center environments,

delivering substantial energy savings. For example, a study on a containerized data center with a micro-grid of cooling resources found that optimizing IT workload scheduling based on ambient conditions could reduce cooling costs by up to 50%, depending on the data center's location and the allowable IT inlet air temperature range. [12]

Two common free cooling methods used in data centers are based on either air-side economizers or water-side economizers.

Air-side economizer

An air-side economizer is a system made up of ducts, dampers, and controls that enables an HVAC system to utilize outdoor air for cooling whenever outdoor conditions are suitable.

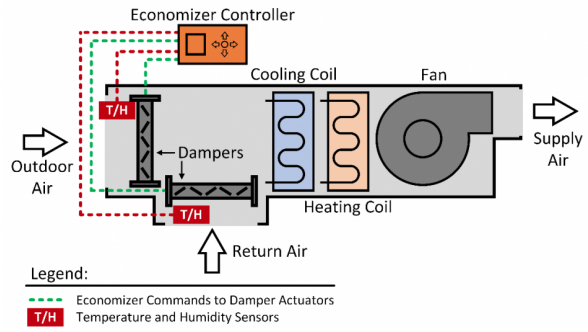


Figure 1.3: Air side economizer. Source: [2]

Unlike systems without an economizer, which depend on a fixed outdoor air damper to admit only a small volume of outdoor air for ventilation and indoor air quality, an economizer adjusts the amount of outdoor air utilized dynamically.

In traditional systems without an economizer, the outdoor air damper is adjusted to provide the minimum required ventilation while optimizing energy use for heating or cooling incoming air, particularly during extreme weather. Even in mild conditions, most of the air passing through the cooling coil is recirculated from within the building, helping to reduce overall energy consumption. In contrast, an air-side economizer boosts energy efficiency by adjusting the supply of outdoor air—ranging from the minimum ventilation levels to 100%—based on the external conditions. When outdoor air can be cooled more effectively than returning interior air, the economizer modifies the intake to optimize the use of outdoor air, thereby decreasing the reliance on mechanical cooling. This adaptive system results in lower overall energy consumption and enhanced cooling performance efficiency. [2]

Water-side economizer

A water-side economizer is a system that enables the direct cooling of chilled water using condenser water when outdoor conditions are sufficiently cold and dry, completely bypassing the mechanical chiller. This method employs a dedicated heat exchanger for heat transfer between the two water loops: the chilled water loop (illustrated in dark blue and yellow in the figure below) and the condenser water loop (depicted in light blue and orange).

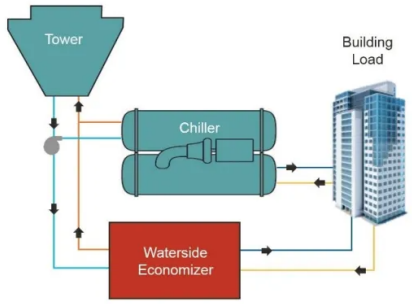


Figure 1.4: Water-side economizer. Source: [3]

By using outdoor conditions to cool the condenser water, which subsequently cools the chilled water, this system removes the need for energy-consuming chiller operation. This technique significantly decreases energy usage, making it a highly efficient and cost-effective cooling solution. [3]

1.3.4 ASHRAE

The primary objective, or the range within which cooling systems must operate to support rooms dedicated to IT equipment, is outlined by several standard organizations. Among these, the most widely adopted is the Thermal Guidelines for Data Processing Environments, developed by ASHRAE Technical Committee 9.9. [4] This guideline offers recommendations for the implementation and maintenance of IT equipment in data centres.

Specifically, ASHRAE's guideline is structured around four key principles:

1. Temperature range
2. Humidity range
3. Classification into four categories (each applicable to data centres)
4. Characteristics of incoming air, including temperature and humidity specifications

Considering these four characteristics, the four categories mentioned earlier are illustrated in Figure 1.5. When the conditions of these environmental classes are met, they ensure the proper functioning of the data center, without exceeding the limits that could harm the equipment. As shown, the recommended envelope, enclosed within Environmental Class A1, lies within a restricted operating range.

It is up to the data center operator to decide whether to operate within this

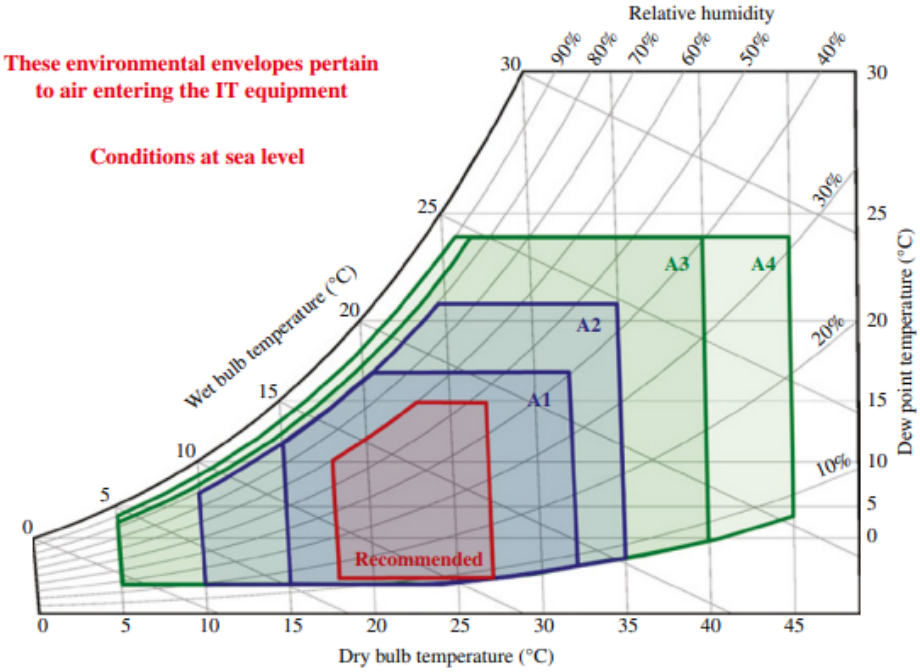


Figure 1.5: ASHRAE Enviromental Classes for Data Centres. Source: [4]

range or extend towards other working classes (A1 excluding the recommended envelope, A2, A3, and A4).

The values upon which these envelopes are based can be found in Figure 1.6.

Equipment Environment Specifications for Air Cooling							
	Product Operation[b,c]					Product Power Off[c,d]	
Class[a]	Dry-Bulb Temperature[e,g], °C	Humidity Range, Noncondensing[h,i,k,l]	Maximum Dew Point[k], °C	Maximum Elevation[e,j,m], m	Maximum Rate of Change[f], °C/h	Dry-Bulb, Temperature, °C	Relative Humidity[k], %
Recommended (Suitable for all four classes; explore data center metrics in this book for conditions outside this range.)							
A1 to A4	18 to 27	–9°C DP to 15°C DP and 60% rh					
Allowable							
A1	15 to 32	–12°C DP and 8% rh to 17°C DP and 80% rh	17	3050	5/20	5 to 45	8 to 80
A2	10 to 35	–12°C DP and 8% rh to 21°C DP and 80% rh	21	3050	5/20	5 to 45	8 to 80
A3	5 to 40	–12°C DP and 8% rh to 24°C DP and 85% rh	24	3050	5/20	5 to 45	8 to 80
A4	5 to 45	–12°C DP and 8% rh to 24°C DP and 90% rh	24	3050	5/20	5 to 45	8 to 80
B	5 to 35	8% to 28°C DP and 80% rh	28	3050	N/A	5 to 45	8 to 80
C	5 to 40	8% to 28°C DP and 80% rh	28	3050	N/A	5 to 45	8 to 80

Figure 1.6: Values delimiting the environmental classes. Source: [4]

1.4 Data centre communication and tasks

A fundamental factor in paving the digital landscape in data centres is the communication between these central hubs of stored, processed, and exchanged data, which is essential to our interconnected world and impacts the performance, reliability, and accessibility of the services and applications relied on a day-to-day basis.

These IT infrastructures can be used for cloud computing and traditional on-premises computing without a cloud model. This chapter and thesis will focus on data centres offering computing resources over the Internet, or, simply put, data centres with cloud computing.

The architecture of cloud computing is multi-layered, with each layer serving a specific purpose.

Physical resources at the base include servers, storage systems, and network equipment. Subsequently, the physical resources are abstracted and then virtualised, allowing for the creation of virtual machines that can be allocated and scaled to meet the changing demands of cloud services.

The orchestration system manages the virtual layer. Its main duties are provisioning, scaling, and balancing workloads across the data centre resources. The orchestration system uses advanced algorithms and machine learning techniques to optimise resource utilisation and minimise latency.

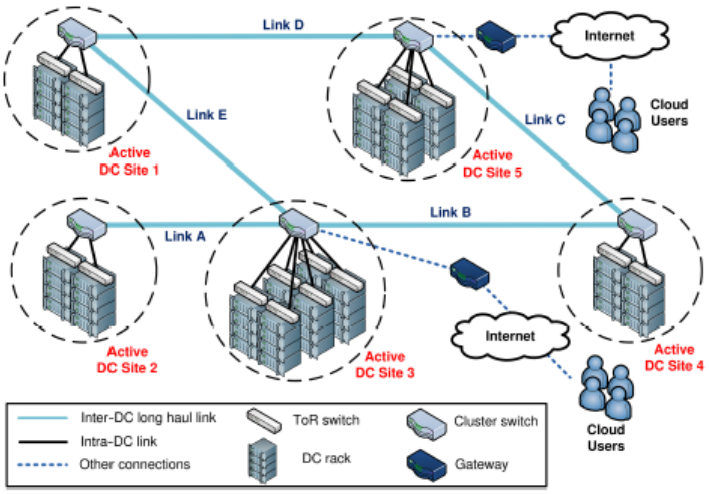


Figure 1.7: Example of geo-distributed Data centre and interconnected sites. Source: [5]

The topmost layer of the cloud computing architecture consists of the cloud services, which can range from simple storage and computing services to more complex functions, such as platform-as-a-service and software-as-a-service.

1.4.1 Communication and techniques utilised in Geo-distributed Data centres

At the simplest level, data centres and end-users communicate through networked connections, where data is transmitted and received using various protocols and technologies. In a simple scenario, an end-user might access a website or application hosted on a local data centre, with the communication occurring over a relatively short distance and low-latency network. However, as the demand for data-intensive services and the need for global accessibility have grown, the landscape of data centre communication has become increasingly complex with the rise of geo-distributed data centre architectures.

One of the key protocols enabling seamless communication between geo-distributed data centres is the Remote Direct Memory Access protocol. This protocol allows for the direct transfer of data between the memory of two computers without involving the central processing unit, reducing communication latency and improving overall efficiency. Another essential protocol is the Web Services Distributed Management protocol, which provides a standardised method for managing and monitoring distributed systems, facilitating the coordination and

control of geo-distributed infrastructure.

Data centres dispersed across multiple geographical regions offer several advantages, such as proximity to end-users, lower latency and better responsiveness, and increased resilience and survivability when faced with failures occurring in loco. [5] Communication, as stated previously, is fundamental. It can add to further challenges, as the data must traverse longer distances over wide-area networks, potentially leading to increased latency and reduced overall performance.

These challenges have been addressed through various strategies and technologies focussed on optimising the communication between geo-distributed data centres and their end-users. One approach is leveraging content on delivery networks, strategically placing caching servers closer to end-users, reducing the distance data must travel and improving response times. [5]

Another method is implementing software-defined networking and network function virtualisation techniques, which allow for dynamic routing and traffic management, establishing efficient routes between data centres and end-users. [5][13]

Additionally, merging technologies such as fog and edge computing, bringing computing resources closer to the point of data generation, can enhance the communication between data centres and end-users by reducing the need for long-distance data transmission. A combination of networking, cloud, and edge computing technology advancements is crucial in maintaining quality of service and service-level agreements. [14]

1.4.2 Differentiating and Assigning Tasks

The distinction between different types of tasks is important in the context of geo-distributed data centres. Tasks can be broadly categorised into "delay-sensitive" and "delay-tolerant". Delay-sensitive tasks, such as real-time analytics or interactive applications, require low latency and immediate response. On the other hand, delay-tolerant tasks can withstand longer processing times, such as batch processing or offline data analysis.[15]

Allocating these tasks within a geo-distributed cloud environment is a critical challenge, as it requires balancing the needs of both types of tasks while also considering factors such as user constraints, security requirements, and available capacity. Techniques such as task partitioning, where workflows are divided and executed at multiple geographic locations, have been proposed to optimise resource utilisation and reduce latency. Homogenous and heterogeneous tasks can be defined as tasks that can't be partitioned, the former, and those that can be partitioned, the latter. [16]

Additionally, using machine learning algorithms to predict task characteristics and allocate resources accordingly has shown promise in improving the overall efficiency of geo-distributed systems.

While the efficient management of tasks in geo-distributed data centres remains an active area of research, the potential benefits are significant. By effectively differentiating and assigning tasks based on their characteristics, cloud providers can deliver improved performance, reliability, and cost-efficiency to their users, ultimately driving the continued evolution of distributed computing.

1.4.3 Quality of Service and Latency

In the context of data centres, quality of service refers to the set of service-level agreements that are established between the cloud service provider and the client. These agreements typically cover parameters such as availability, throughput, response time, and latency. [17]

Guaranteeing the quality of service can be challenging, particularly in complex service networks composed of multiple service components. [17]One key factor that can impact the quality of service is latency, which refers to the time delay experienced by data as it travels from the user to the data centre and back.

Various factors, including network congestion, the distance between the user and the data centre, and the availability of resources within the data centre, can influence latency.

The distance between data centres and the user influences the latency. This is because the time it takes for data to travel from the user to the data centre and back is directly proportional to the physical distance between them. As the distance between the user and the data centre increases, the latency experienced by the user also tends to increase. This is a critical factor to consider when designing and deploying geo-distributed data centres, as the placement of these data centres can significantly impact the latency experienced by users.

For example, a user in Europe accessing a data centre in the United States may experience higher latency than one in the same region as the data centre. The threshold of acceptable latency can vary depending on the specific service-level agreements and the requirements of the application or service being provided.

The increase in latency with distance can be significant. For example, studies have shown that a circa 161 km increase in distance between a user and a data centre can result in an additional 5-10 milliseconds of latency [18]. Similarly, a 1610 km increase in distance can lead to an additional 50-100 milliseconds of latency. Furthermore, the threshold of acceptable latency can vary widely depending on the application requirements. For real-time, latency-sensitive applications such as video conferencing or online gaming, the latency threshold is

typically as low as 10 milliseconds. In contrast, less time-critical applications like email or file transfers may have acceptable latency thresholds of 100 milliseconds or more.

As mentioned above, deploying geo-distributed data centres can provide users with greater proximity to the data centres and higher survivability in the event of failures. By strategically placing data centres in different geographical regions, cloud providers can offer their clients lower communication latency and better services.

Additional complexities can, though, be introduced, such as the need to manage data replication and synchronization across multiple sites. Optimising the trade-off between latency and survivability is a critical challenge for cloud providers as they strive to meet their clients' evolving demands while maintaining the efficiency and cost-effectiveness of their infrastructure. [18]

Chapter 2

Simulation tool and study cases

A data centre can be seen as a black box requiring two energy vectors: electricity and cooling. Electricity is essential for powering the main components that offer a service, specifically the racks of servers, ensuring they are online and functioning while also feeding additional loads such as UPS and lighting. Cooling, which is initially derived from electricity but later transformed, guarantees that the server rooms maintain adequate conditions for the continuous operation of the IT equipment.

In any energy system, energy vectors may originate from various sources. In the context of data centres, electricity can be sourced from the grid or onsite generation, such as photovoltaic (PV) plants. Similarly, the cooling provided for IT equipment may come from a refrigeration system or, if conditions permit, from the external environment, utilising the free cooling technique.

2.1 Introduction to XEMS13

XEMS13, an optimisation tool developed by the Energy Department of Politecnico di Torino "Galileo Ferraris" and LINKS, is capable of simulating polygeneration systems and optimising their management, with the primary objective of minimising operational costs while considering both technical and operational constraints. [6]
This optimisation tool can be employed to evaluate the advantages of free cooling and outline the benefits it brings to such applications.

Generally, XEMS13 requires various inputs; as these inputs are processed, a corresponding amount of outputs will be produced. The required inputs are as follows:

1. Profiles, which will include the following:
   - The load trends of the systems under consideration, which in this case are electricity and cooling of the locations assessed;
   - The trends of electricity prices of the locations taken into consideration;
   - Temperature and relative humidity trends of the locations taken under consideration.
2. Netlist
3. An ".xml" file which are the components, in this case of the data centres undergone in the study

The outputs include:

1. A ".csv" file, results of optimised electrical and cooling energy consumption both
2. A ".xml" file that will give us an overall sum of the consumption of energy (this will depend on the trends of electricity prices)
3. A ".txt" file (useful for the creation of Sankey diagrams)

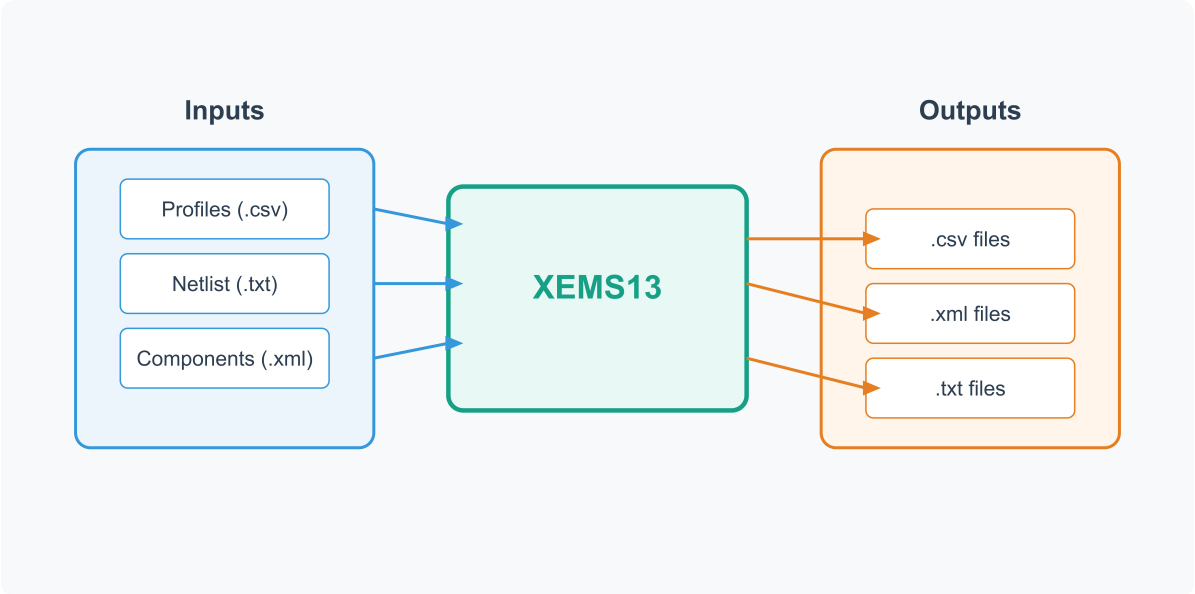


Figure 2.1: Data flow of XEMS13

Inputs

For the specific situation at hand, the first input will be the profile trends of our case study data center, based on data monitored over one year. To simplify the analysis, the observed year will be broken down by taking one-hour samples over the course of one week per month, for a total of 12 weeks, effectively simulating one year. This approach will be applied to each individual profile used in the study.

The acceptable operating conditions will be determined by external environmental factors (specifically ambient temperature, relative humidity) and the cooling demand of the system on an hourly basis, which defines the boundary within which the system components can operate. If operation within this boundary is not possible, the simulation will yield an infeasible or unbounded result. In other words, XEMS13 will fail to find an optimal solution and determine that the scenario is not viable.

The load and electrical price trends file consists of ".csv" files containing data across 168 rows. As mentioned above, we will sample on an hourly basis for the week in question, leading to 168 hours for the entire week. There will be six columns, with the last column containing data related to loads or electrical prices. The other rows represent the year, month, week, day, and hour to which the particular data pertains.

Since we sample hourly, filling in this information will not be necessary. The types of loads we will be addressing are electrical and cooling loads. The cooling loads will be derived from our electrical loads, as the cooling load will be generated through a chiller powered by electricity. It is crucial to input the electrical load excluding the electrical load required to generate the cooling load; the latter will be entered separately as the cooling load. The loads are inserted as kW, and as this is done on an hourly basis, this will then match the energies or kWh.

The electrical prices are also ".csv" files and can be split into two different types: Cp, or electricity purchased from the grid [€/Mwh], and Cs, or electricity sold to the grid [€/Mwh]. For the case at hand, only electricity purchased from the grid will be evaluated, as data centres, being highly energy-consuming, will never find the occasion to sell back to the grid. The ".csv" files with information to fill out follow the same logic as the load trends. Other trends that will be useful in evaluating the potential use of free cooling are the outside temperature and relative humidity. These are also inserted as ".csv" files.

The temperature and relative humidity trends are also provided as .csv files to be used as input data for the analysis.

The second input, the .xml file or the components file, as its name suggests, provides a detailed list of the components that make up the analyzed systems.

Each component is defined by various characteristics.

This ".xml" file containing all the components of our "plant," which is morphed for this specific case into a data centre. The components responsible for electrical supply that make up a generic system would be the electrical grid, onsite generation (such as CHPs) or PV arrays. The cooling components can be composed of refrigerator groups in generic form, absorption chillers and refrigerator groups with cooling towers.

In our specific case, the system under study being a data centre, the electrical supply will be provided exclusively by the local grid. The cooling components will consist of a compression chiller accessubg the cooling tower through the compressor unit (the only component operating when free cooling is unavailable) and the compressor chiller bypassing the compressor unit to the cooling tower (activated when free cooling is available).

The information for the grid that interests the case study at hand will be the emission factor of the grid from which electricity is drawn; other information is included in XEMS13 but is useful in the case a load flow study were to be performed on the distribution level.

Whereas the systems responsible for cooling our plant typically has their rated power (measured in kW), their COP, and a name used to distinguish each type of cooling plant from another. Input powers and output powers are inserted to have XEMS13 understand the characteristic curve of the cooling component. The cooling components' behaviour is not linear. XEMS13 will discretise the component's characteristic curve, approximating the non-linear behaviour with a linear behaviour. For non-standard refrigeration units, further specifications are required, such as an absorption chiller, these would be including the nominal exit temperature and the COP variation coefficient (a coefficient that adjusts your COP based on the exit temperature). This will also contribute to the definition of a boundary within our system can work.

The third input, the netlist, is a .txt file containing a set of tags essential for the simulation. Specific components with defined characteristics will be listed in this file based on the desired simulation conditions. As expected, these tags reference the data provided in the two previous inputs (profiles and components). For example, a netlist corresponding to a specific week will include the tags of all previously mentioned profiles combined with the components. This defines a well-structured boundary within which an optimal solution can be sought.

Outputs

The first useful output file is a ".csv" file, which will produce the scheduling of resources divided by energy vector. The tendency of what source our hypothetical data centre will employ is given. In this case, two energy vectors will be scheduled: ELE, representing electrical energy, and COL, representing cooling energy. In other words, it will give us a description of where our system will get its electrical energy if multiple sources are available, that being from the grid or any other onsite generation. Likewise, for cooling, a distinction on whether free cooling was possibly tapped into (implying energy savings for the data centre) and for how long will be given; otherwise, the cooling will be provided by the cooling plant.

The ".xml" output file will encompass the complete results of the optimisation process. Control variables will be introduced, with a numeric value assigned to each. These control variables are determined when selecting the strategy for optimisation. The possible strategies are: ECO, ENVI, and ECOENVI. These three approaches represent the minimisation of costs (ECO), minimisation of emissions (ENVI), and minimisation of costs as the prime objective while retaining emissions as a significant factor (ECOENVI).

The third output is a .txt file that will be useful in the creation of Sankey diagrams.

2.1.1 How the cooling component is modeled on XEMS13

A more detailed explanation of how the cooling component is modeled in the optimization tool will be provided, as it plays a key role in achieving further optimization for the case at hand.

The cooling system modeled in our specific case is an electric compression chiller. The diagram below provides an overview of how this type of chiller functions in real world application. As shown on the left side of Figure 2.2, the electric chiller works by extracting heat from the process load and transferring it to a refrigerant circulating within the evaporator, facilitated by the chilled water loop. An electrically powered compressor then compresses the refrigerant vapor, which increases its temperature due to the heat of compression. The accumulated heat is then passed into the condenser water loop, ultimately discharged into the atmosphere via a cooling tower. A simplified free cooling (FC) configuration proposed for the plant is depicted on the right side of Figure 2.2. This system incorporates a bypass arrangement, which uses three-way valves along with extra piping to link the condenser water loop directly to the chilled water loop, creating a unified circuit. This design enables heat from the process load to be directly expelled into the atmosphere through the cooling tower, eliminating the need for the chiller compressor.

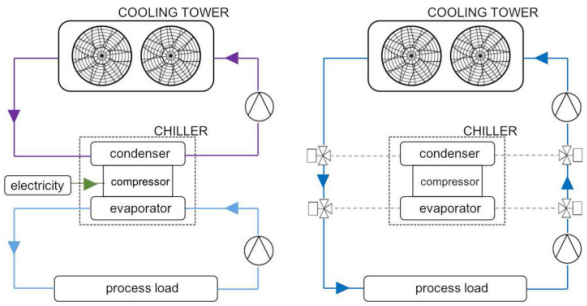


Figure 2.2: Refrigerant Compression Chiller Systems. Left side: Normal Operation. Right side: Free Cooling operation. Source: [6]

Clearly, during free cooling (FC) operation, when the bypass is engaged, the associated chiller cannot operate and must remain turned off.
The discharge temperature used in both scenarios, whether free cooling is available or not, will be set at 10°C. This is an acceptable level, which may result in slightly warmer conditions in the data hall compared to the lower limit of 7°C. However, studies have shown that maintaining temperatures within the 7°C to 12°C range ensures the data hall remains within acceptable operating conditions.[7]. The return temperature will be set at 27°C.

Free cooling is coded as though the component has an infinite COP, which means that with a very small amount of power input into the fictitious component, a significantly larger amount of cooling power would be produced. This corresponds to reality because no power is actually required, and the cooling power is delivered without any actual "input". To keep XEMS13 in check so that the free cooling function is not regularly turned on and off for unreal amounts of time, in other words to make sure the simulations are within the realm of the doable, two other values have been assigned to this component: MOT (Minimum on Time) and MOS (Minimum Shutdown Time). These two values assure that the component acting as our "free cooling machine" if turned on will at least be on for minimum 2 hours, otherwise this option will not be activated.

The cooling towers can be of three different types: vapor (water based like liquid cooling towers), air (air based like economizers) and geo (geothermic like heat sinks). In the current study only vapor will be considered as the means of cooling.

Concluded the explanation of the inputs and interworkings of XEMS13 the Netlist will make more sense as it recalls all of these components and lists them as a set, from the components constituiting our system, the scheduling, the type of optimisation chosen, the system voltage and dispactchable electric and thermic inputs.

2.2 XEMS13 applied to our study cases

Four scenarios will be created and combined to assess the benefits and differences each scenario offers compared between each other. We can categorize these scenarios into macro and micro scenarios. Micro scenarios represent the operation of a single data center in isolation, while macro scenarios depict the coordinated operation of multiple data centers functioning as a unified system.

The naming convention for the micro scenarios will be as follows:

- **Base Scenario**: The data centre in this scenario is equipped with a cooling plant that is not designed for free cooling but is capable of meeting the facility's cooling demand;
- **Free Cooling Scenario**: This scenario is equivalent to the base case, with the added capability of satisfying the cooling load through free cooling whenever environmental conditions are favorable;

The naming convention for the macro scenarios will be as follows:

- **Load Management Scenario — Approach 1 (Scenario A1)**: Two datacentres in geodistributed locations will be considered, in this study they will be Barcelona and Dublin. They will be identical in size and equipped with the same cooling systems. The two geographically separated data centers are interconnected through cloud services, allowing for load management and distribution, shifting load from Barcelona to Dublin. This interconnection enhances operational flexibility and efficiency
- **Load Management Scenario — Approach 2 (Scenario A2)**: This scenario closely resembles Scenario A1, with the addition of an extra assumption regarding load management.

For the sake of brevity, the Dublin and Barcelona data centres will hereafter be referred to as D and B, respectively.

The base scenario will serve as our starting point, representing the situation that is to be optimized whether it is applied to B or D.

For each scenario, the year under consideration for the electricity price profiles will be 2023. To represent monthly trends, the first full week of each month spanning from the first Monday to the first Sunday will be selected. This sampling method will be applied consistently across all profiles, ensuring that each dataset covers the same period. The approach is designed to provide a representative snapshot of each month's conditions.

The choice of 2023 is justified by the fact that it is the most recent year with a complete set of relevant data available. Ideally, it would have been preferable to select an additional pre-pandemic year unaffected by the COVID-19

pandemic and the Russia-Ukraine war, as such a year would have provided a broader perspective on the potential advantages offered by the alternative scenarios compared to the base scenario (i.e electricity prices not being impacted by social and geo-political factors).

The parameters that will vary across the scenarios are the cooling load and the electric load. All remaining profile inputs will remain consistent throughout all scenarios.

2.2.1 Base scenario

Load trends - Cooling load

The data centre considered in this study has a capacity of 1 MW. All subsequent calculations and analyses in each of the four scenarios will be based on this specified capacity.

As indicated by several IT server manufacturers, including ABB [19], the idle power consumption of servers typically accounts for approximately 30% of the power drawn under maximum utilization. In other words, a baseline power demand of 30% is consistently present. Considering this and assuming both data centres (applicable to both previously mentioned locations, Dublin and Barcelona (in the load management scenario)) primarily operates during standard working hours, the following load profile can be established in Figure 2.3.

As observed, the minimum load experienced by these servers corresponds to 50%, which consists of approximately 30% attributed to idle operation and 20% associated with active task processing and service delivery. The maximum load is capped at 90%, leaving a 10% buffer capacity to accommodate any additional processing requirements. This load profile accurately reflects the typical operational behavior of both data centres during a standard day.

Peak hours are defined as periods when the server utilization percentage is greater than or equal to 80%, occurring between 9:00 a.m. and 6:00 p.m. (09:00 to 18:00). A typical operational day begins with an initial utilization of 50%, which gradually increases until reaching the maximum utilization level. The peak utilization of 90% is maintained between 11:00 a.m. and 5:00 p.m. (11:00 to 17:00).

In total the utilisation factor of the data centre(s) throughout the year will equate to:

$$UF = \frac{Hours_{D\,and\,B}}{Hours_{Year}} = \frac{6186.75\ hours}{8760\ hours} = 0.70625\ *100 = 71\% \qquad (2.1)$$

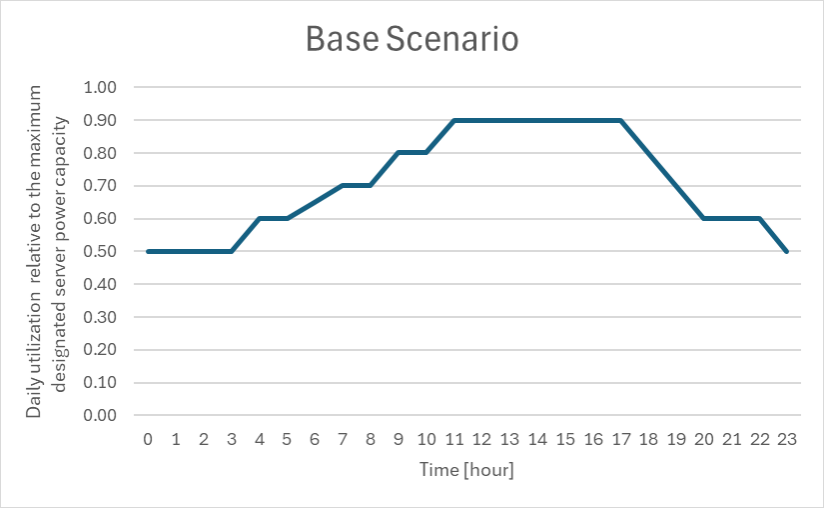


Figure 2.3: Daily Utilization in the Base Scenario Relative to Maximum Designated Server Power Capacity

The distribution of power demand across the two identical data centres is illustrated in Figure 2.4. As expected, the majority of the power demand is attributed to the IT equipment, with cooling systems representing the second largest energy consumer. Of the total 27.5% of power allocated to cooling, approximately 90% is directly used to cool the IT equipment (3 capitol free cooling), while the remaining 10% is allocated to auxiliary systems.

For the purpose of this study, the focus will be on optimizing 25% of the total power demand, primarily related to cooling operations. The remaining 20.5% of the total power demand is allocated to lighting, power distribution within the data centre infrastructure (including transformers, switches, and uninterruptible power supplies (UPSs)), as well as air distribution within both the data halls and office areas.

The breakdown shown in Figure 2.4 will give us a PUE (power usage effectiveness) of 1.92, meaning that for every 1.92 W drawn 1 W will be designated to the IT equipment. Aligning with the international average established in 2016 in Figure 3.2.

Since 25% of the total power demand is allocated to the cooling of IT equipment [10], and given that each data centre has a total capacity of 1 MW, the cooling load for the data halls can be calculated by multiplying the total data centre capacity by this percentage.

Several considerations were made to calculate the electrical load to be input into XEMS13. The initial assumption was that, at every instant, there would

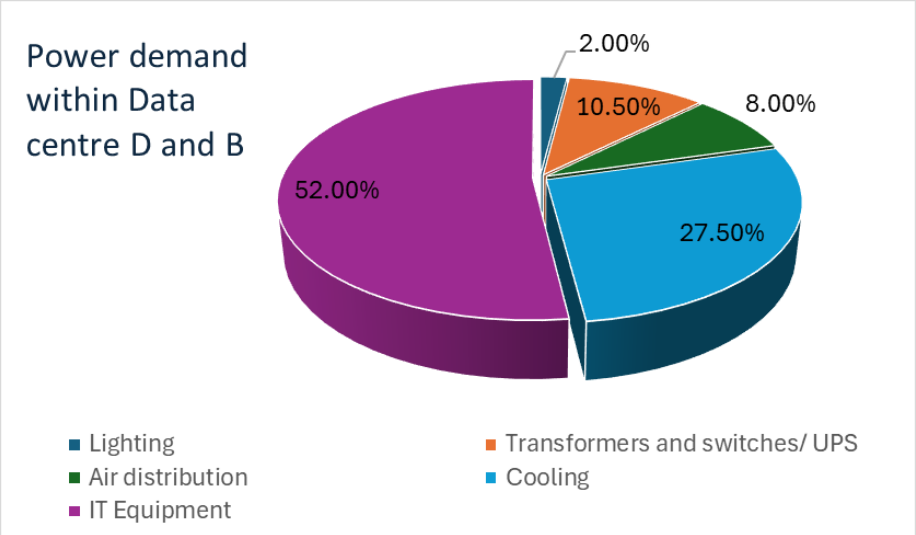


Figure 2.4: Power demand ripartition within D and B data centres

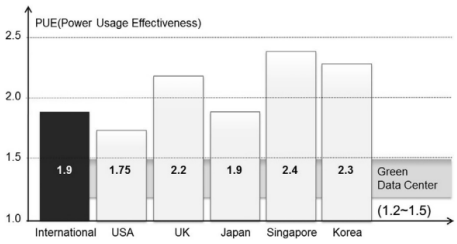


Figure 2.5: Comparison of Power Usage Effectiveness (PUE) Across Global Data Centres. Source: [7]

be a load demanding a constant amount of power. This, when looking at the power demand breakdown, translates to air distribution, transformers, switches and UPSs, lighting, and the overall cooling, with 10% allocated to non-IT equipment. In total, this amounts to 23% of the overall power demand. The variable parts of the power demand, which depend on the functioning of the servers, include 52% allocated to IT equipment and 90% of the cooling, which accounts for 25%.

$$P_{ele} = (0.23 \cdot P_{SizeDC} + K \cdot 0.52 \cdot P_{SizeDC}) + K \cdot 0.25 \cdot P_{SizeDC} \qquad (2.2)$$

The calculation of the electrical power load on an hourly basis can be seen in 2.2. Anything in parentheses in 2.2 will be used for the calculation of electrical loads, while the remainder will be allocated to the calculation of cooling load. K represents the IT equipment utilization, which varies from hour to hour. In

our case, this value can range from 0.4 to 1.This calculation will be performed for each scenario. The varying factor across scenarios will be the hourly values of K, representing the IT equipment utilisation.

From the calculated power values, the corresponding energy values (MWh) can then be derived using the following equation:

$$E_i = \int_{t_i}^{t_{i+1}} \left(P_i + \frac{P_{i+1} - P_i}{\Delta t}(t - t_i) \right) dt \tag{2.3}$$

$$E_i = \left[P_i t + \frac{P_{i+1} - P_i}{\Delta t} \frac{(t - t_i)^2}{2} \right]_{t_i}^{t_{i+1}} \tag{2.4}$$

$$t_i \, and \, t_{i+1} = t_i + 1 : \tag{2.5}$$

$$E_i = P_i(t_{i+1} - t_i) + \frac{P_{i+1} - P_i}{2}(t_{i+1} - t_i) \tag{2.6}$$

$$E_i = P_i + \frac{P_{i+1} - P_i}{2} \tag{2.7}$$

$$E_i = \frac{P_i + P_{i+1}}{2} \tag{2.8}$$

Once the energy values have been calculated, they are further adjusted by applying the hypothesized Coefficient of Performance (COP) of the cooling plant, which is assumed to be 3.5. This COP is throughout all scenarios.

Upon completing these calculations, it can be observed that during a typical day, the minimum energy demand of the cooling plant amounts to approximately 125 kWh, while the maximum energy demand reaches 225 kWh. Whereas the cooling requested to keep the datahall room under acceptable conditions has a minimum of 437.5 kWh and a maximum of 787.5 kWh.

The calculations performed for the base scenario will also apply to the subsequent scenarios, with the only differences being that the cooling demand may be partially met through free cooling when favorable environmental conditions occur, and the load profile may be altered due to the application of load management strategies.

The load profile for a typical day will be repeated consistently throughout each week, with no variations. To simulate the behavior over an entire year, one representative week per month will be selected, resulting in a total of 12 weeks used to approximate the annual operational profile.

This approach will be applied in conjunction with all relevant yearly data, including ambient temperature, relative humidity, electric load, and trends in electricity prices, both for purchased electricity and electricity sold to the grid.

Load trends - Electric load

A similar approach to the one used for deriving the cooling load will be applied to calculate the electric load. The electric load of interest refers to the net electric load, excluding the portion allocated to cooling. Referring to Figure 2.4, this corresponds to all power demand not directed towards the cooling systems serving the data halls.

This portion accounts for 75% of the total power demand, which is three times the amount allocated to the cooling of IT equipment.

The maximum electricity consumption is observed at 675 kWh, while the minimum electricity consumption is approximately 375 kWh.

Temperatures and Relative Humidity

The PVGIS platform directly provides the temperature and relative humidity trends. Given the geographical differences between the two locations, a Typical Meteorological Year (TMY) will be used to evaluate the climatic conditions for both data centre B and data centre D.

The Typical Meteorological Year for B is as follows: Table 2.1, while the Typical Meteorological Year for D is as follows in Table 2.2. The temperature and relative humidity profiles used in this study will correspond to these representative weather datasets.

The Typical Meteorological Years described above will also be applied across the remaining scenarios. Consequently, the weather data and conditions outlined will be consistently used throughout the various simulations.

Month	Year
Januaray	2008
Feburary	2006
March	2007
April	2015
May	2007
June	2007
July	2007
August	2011
September	2017
October	2016
November	2022
December	2018

Table 2.1: TMY of B

Month	Year
Januaray	2015
Feburary	2008
March	2013
April	2020
May	2007
June	2010
July	2012
August	2007
September	2013
October	2015
November	2015
December	2014

Table 2.2: TMY of D

Price trends - Sold and purchased electricity prices

Precise electricity price trends were not directly available; therefore, they were derived using data from the day-ahead market marginal prices provided by the respective market operators—SEMO for Dublin (D) and OMIE for Barcelona (B)—along with the average semesterly prices for non-household consumers.

The Eurostat database was used to obtain the average electricity prices for the two semesters of 2023. The Table 2.3 presents the average price for each semester, as well as the overall average price for the entire year.

Country	Average S1 [€/MWh]	Average S2 [€/MWh]	Average 2023 [€/MWh]
Ireland	307	279.8	293.4
Spain	191.2	183.7	187.45

Table 2.3: Semesterly Averages and Overall Yearly Average (2023) for Data Centres B and D

To estimate the hourly electricity prices for 2023, the average price obtained from Eurostat will be compared against the average marginal prices provided by the respective market operators. Specifically, the difference between the Eurostat average price and the average marginal price will be calculated. This differential value will then be added to the hourly day-ahead marginal prices, thereby adjusting the hourly prices to align with the semesterly averages reported by Eurostat.

The Tabel 2.4 presents both the average marginal prices and the calculated adjustment values (i.e., the difference between the average 2023 electricity

price and the average marginal price) for Dublin (D) and Barcelona (B).

Country	Average Marginale Price [€/MWh]	Avg.€/MWh - Avg.Marginal Price [€/MWh]
Ireland	121.91	171.49
Spain	90.71	96.736

Table 2.4: Semesterly Averages and Overall Yearly Average (2023) for Data Centres B and D

The purchased price (C_p) will be calculated by adding the adjustment value (as shown in Table 2.4) to the hourly day-ahead market prices provided by the respective market operators. Conversely, the selling price (C_s) will correspond directly to the actual hourly day-ahead market prices supplied by the market operators.

The purchased price (C_p) and selling price (C_s) for both data centre D and data centre B will remain consistent across all scenarios.

2.2.2 Free cooling scenario

As previously mentioned, this scenario is identical to the base scenario, with the sole addition being the ability to utilise free cooling when favourable environmental conditions are met. In this case, the cooling load and electric load, which have already been established as variable parameters across scenarios, will initially follow the same profile as the base scenario.

All assumptions made in the base scenario remain valid and applicable in this scenario.

2.2.3 Load Management Scenario - Approach 1 (Scenario A1)

The interconnection of cloud services between data centres D and B introduces significant opportunities for load management optimisation. Tasks from users near Data Center B may be redirected to Data Center D, leveraging D's free cooling potential by shifting the load from B to D.

Due to its geographical location, data centre D has a higher probability of experiencing favourable conditions for free cooling, with lower average temperatures recorded throughout the year (9.6°C compared to 16.2°C). This

factor explains why, in this scenario, and as will be demonstrated in the following subsection (as will be shown in scenario A2), workload will be shifted in a scheduled manner from data centre B to data centre D whenever conditions allow, for a to be confirmed and supposed more energy-efficient operation.

In this scenario, load management strategies will be implemented to optimise the combined energy consumption of both data centres, which will be treated not as two independent systems, but rather as components of a single, integrated system.

The load shifted from B to D results as 6.86% of the total cooling load (cooling load of B summed with that of D).

Load trends - Electric and Cooling load

Using the cooling and electric loads established in the base scenario, a load-shifting schedule will be developed to partially transfer the load from data centre B to data centre D. A simple load management strategy will be applied, based on the assumption that idle server operation corresponds to 30% of maximum power capacity. Any load exceeding this 30% threshold will be evenly distributed between data centres B and D.

This rule will apply consistently, with the exception of periods where server utilization already reaches 90% of the maximum designated power capacity — from this point onward will be referred to as the "buffer region". In these cases, further load redistribution will not occur to preserve operational stability. Applying these hypotheses results in the following utilisation factors (UF):

$$UF_{D_{A1}} = \frac{Hours_{D_{A1}}}{Hours_{Year}} = \frac{7035.38\ hours}{8760\ hours} = 0.803\ *100 = 80.3\% \tag{2.9}$$

$$UF_{B_{A1}} = \frac{Hours_{B_{A1}}}{Hours_{Year}} = \frac{5338.13\ hours}{8760\ hours} = 0.609\ *100 = 60.9\% \tag{2.10}$$

At first glance, data centre D experiences an increase in utilization of approximately 9.3%, while data centre B sees a decrease of approximately 10.1%. The resulting load profiles are presented in Figure 2.6 and Figure 2.7.
During off-peak hours, the load at data centre B has been reduced, reaching a minimum utilization of 40%, compared to the 50% observed in the base scenario. Conversely, data centre D experiences an increase in minimum utilization, reaching 60%. Overall, this results in a more balanced and flattened utilization profile across the latter mentioned data centre.

With this load management approach, the buffer limit of 90% is reached earlier, specifically at 7 a.m., compared to 11 a.m. in the base scenario. This

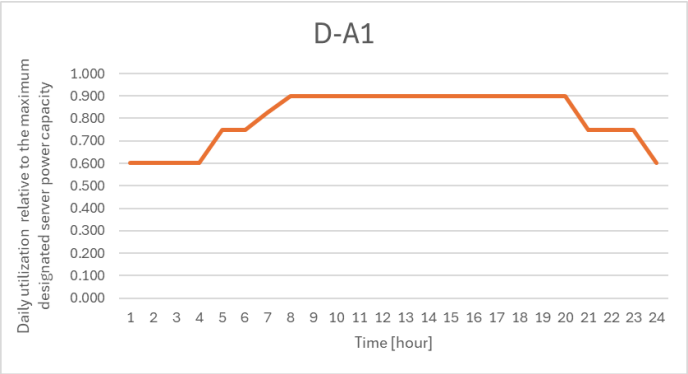


Figure 2.6: Daily Utilization in the A1 Scenario of D Relative to Maximum Designated Server Power Capacity

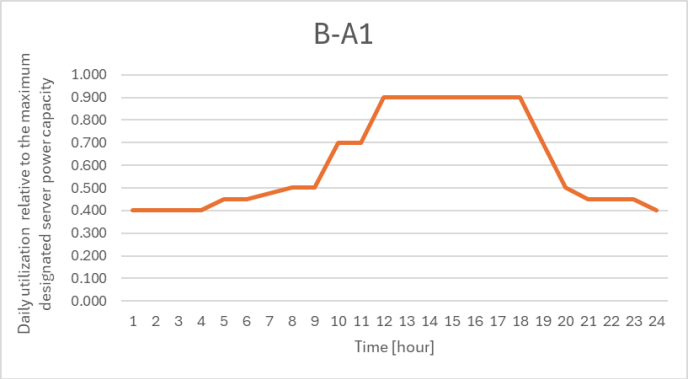


Figure 2.7: Daily Utilization in the A1 Scenario of B Relative to Maximum Designated Server Power Capacity

results in the servers at data centre D operating at maximum designated capacity for a longer period of time. Meanwhile, at data centre B, the duration during which servers operate within the buffer region remains unchanged.

The maximum electricity consumption (excluding cooling) for data centre D is 698 kWh, while the minimum consumption is 542 kWh. The cooling load to be satisfied ranges from a maximum of 787.5 kWh to a minimum of 525 kWh.
In comparison, data centre B will experience a maximum electricity consumption (excluding cooling) of 698 kWh, identical to data centre D, but its minimum electricity consumption will drop to 438 kWh. The cooling load for data centre B will match data centre D at the maximum value of 787.5 kWh, while the minimum cooling load will be 350 kWh.

In both the base scenario and the free cooling scenario, if the B and D data centers are considered as a whole, the cooling load is evenly split (50% is cooled by B and 50% by D). With this approach, 6.85% of the cooling load (which will also equate to the electrical load) is shifted from B to D.

As previously mentioned, the electricity price trends (both purchased and sold), along with temperature and relative humidity data, will remain consistent with those used in the base scenario.

2.2.4 Load Management Scenario - Approach 2 (Scenario A2)

In this fourth and final scenario, an additional consideration is introduced to build upon the framework established in Scenario A1. Specifically, the "buffer region" and the 10% buffer limit will no longer be applied, thereby allowing the servers to operate at their full designated power capacity of 100%.

Following the same reasoning and considerations applied in Scenario A1, this adjustment will result in a slightly greater amount of load being shifted from data centre B to data centre D. The amount of cooling load moved from B to D results as 9.81% of the total cooling load of B and D.

Load trends - Electric and Cooling load

There is a slight increase in the utilization factor (UF) for data centre D, with a corresponding decrease in the utilization factor for data centre B.
newline

$$UF_{D_{A2}} = \frac{Hours_{D_{A2}}}{Hours_{Year}} = \frac{7400.38\ hours}{8760\ hours} = 0.845\ *100 = 84.5\% \tag{2.11}$$

$$UF_{B_{A2}} = \frac{Hours_{B_{A2}}}{Hours_{Year}} = \frac{4973.13\ hours}{8760\ hours} = 0.568\ *100 = 56.8\% \tag{2.12}$$

The utilization trend for data centre D shows that 100% of the designated server capacity is reached between 10:00 a.m. and 7:00 p.m. (10:00 to 19:00). As observed in the A1 approach, the minimum utilization remains at 60% throughout. Alternatively, data centre B reaches a maximum utilization of 80%, maintained from 11:00 a.m. to 5:00 p.m. (11:00 to 17:00). The maximum electricity consumption (net of that consumed for cooling) in data centre D under the A2 approach — which represents the highest value across all four scenarios — reaches 750 kWh, while the minimum, at 542 kWh, remains the same as in the A1 approach. The cooling load ranges from a maximum of 875 kWh to a minimum of 525 kWh.

In contrast, data centre B experiences a peak hour utilization corresponding to

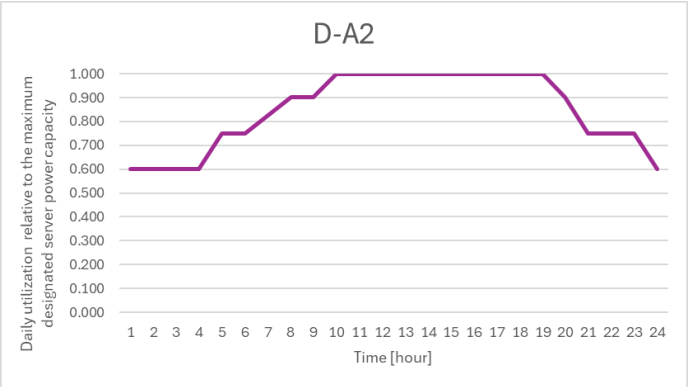


Figure 2.8: Daily Utilization in the A2 Scenario of D Relative to Maximum Designated Server Power Capacity

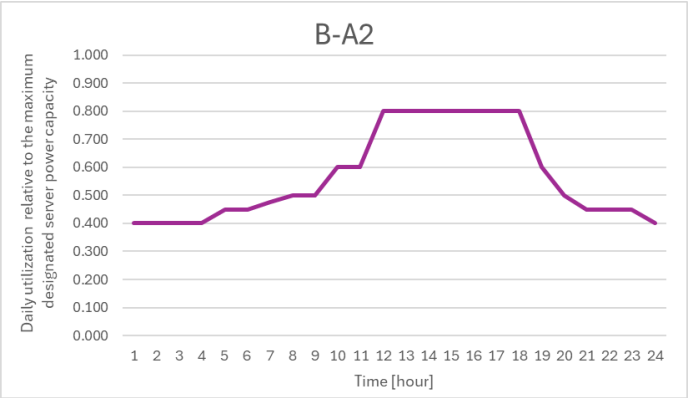


Figure 2.9: Daily Utilization in the A2 Scenario of B Relative to Maximum Designated Server Power Capacity

an electric load (net of cooling) with a maximum of 646 kWh and a minimum of 438 kWh. The cooling load for B ranges from a maximum of 700 kWh to a minimum of 350 kWh.

As mentioned when making considerations in Approach 1, if the B and D data centers are considered as a whole, the cooling load is evenly split (50% is cooled by B and 50% by D). With this approach, 9.86% of the cooling load (which will equate to the cooling load) is shifted from B to D As noted earlier, the trends for electricity prices — both for purchasing and selling — as well as temperature and relative humidity data, will be kept the same as in the base scenario.

Chapter 3

Analysis of Results

A total of four study cases will be analysed in this chapter. Each case study will specify the number of data centers considered (either only D or both B and D) and the corresponding scenarios under which the data center(s) operate. The cases analysed through the optimisation tool are in total:

1. **Study Case 1**: This case evaluates the benefits of Data Center D operating with free cooling capabilities compared to operating without them. Specifically, it compares D's performance under the base scenario versus the free cooling scenario.

2. **Study Case 2**: This case utilizes the A1 macro scenario, where load management is established between Data Centers B and D. The micro scenario applied to B corresponds to the base scenario, while D operates under the free cooling scenario. In other words, B, which does not have access to free cooling, shifts a portion of its load to D, which benefits from free cooling.

3. **Study Case 3**: This case is similar to Study Case 2 but differs in the macro scenario, which changes from A1 to A2. In this scenario, both Data Centers B and D operate under the same micro scenario. Additionally, a slightly larger portion of the load is shifted from B to D.

4. **Study Case 4**: In this case, the macro scenario remains A2, and both Data Centers B and D operate under the free cooling micro scenario. This scenario represents an extreme case, assessing whether load management remains beneficial when both locations have access to free cooling.

To avoid infeasible or unbounded solutions, the chiller size will be set to 1600 kW, the minimum capacity that meets this requirement. Consequently, the cooling tower will be sized accordingly.

Except for Study Case 1 all considerations of numbers dealt will be of the sampled weeks.

3.1 Study Case 1

Before making any comparisons, it is important to first understand Data Center D's energy consumption under the assumption free cooling is not a solution. The total energy consumption of the data center, accounting for all components represented in the pie chart in Figure 2.4, is 1.568 GWh. Under the conditions set in the base scenario, 23.22% of this energy is used for cooling, which amounts to 0.364 GWh.

These figures are based solely on the sampled weeks. If we extrapolate by assuming a consistent monthly energy consumption pattern, the total annual consumption would be 6.813 GWh, with 1.582 GWh allocated to cooling the IT equipment.

The simulation results, shown in Figure 3.1, illustrate the hours within each week of the selected months that Free Cooling was utilised. The blue-marked sections indicate when free cooling was utilized, while the red-marked sections indicate when it was not. The months when free cooling was un-

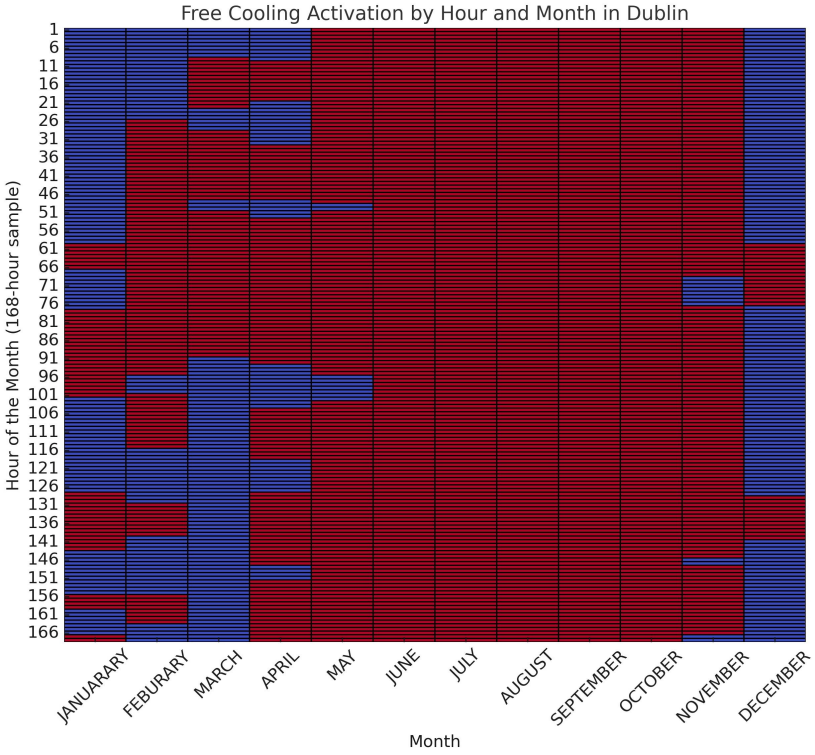


Figure 3.1: A visual understanding of when Free Cooling is tapped into

available span from June to October, whereas from November to May, it was successfully utilized. Although the sampling method considered the first Monday to the first Sunday of each month, the results align well with average temperature and relative humidity trends. However, November may have the potential to offer more opportunities for free cooling. In total, 487 hours of free

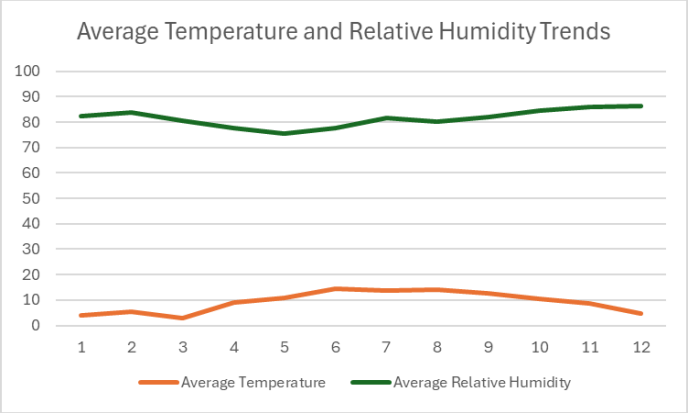


Figure 3.2: Average temperature and relative humidity trends of Dublin

cooling were utilized out of the 2,016 hours sampled. December recorded the highest usage, with 139 hours of free cooling access. Table 3.1 summarizes the results, detailing the percentage of each month's cooling demand met by free cooling and the proportion of total cooling for each month that was satisfied through free cooling. A significant 22.5% of IT cooling demand savings is

Table 3.1: Overview of access to free cooling in D datacentre

Month	$E_{ChillFC}$ **[kWh]**	**% of Total Cooling Demand**	**Hours of FC**
1	68 526	66.01%	115
2	36 715	35.36%	66
3	55 796	53.74%	95
4	25 557	24.62%	51
5	4181	4.03%	9
6	0	0.00%	0
7	0	0.00%	0
8	0	0.00%	0
9	0	0.00%	0
10	0	0.00%	0
11	5558	5.35%	12
12	84 581	81.47%	139
OVERALL	28 0914	22.55%	487

achieved solely by implementing two three-way valves to bypass the compressor

unit, leading to a corresponding cost reduction (the figure stated above does not account for the energy spent for the cooling towers). Table 3.2 lists the months with access to free cooling, comparing their costs with and without free cooling. Additionally, the percentage reduction in costs when transitioning from non-free cooling to free cooling is provided. This results in an overall

Table 3.2: Reduction in costs with the implementation of Free cooling

Month	**Costs with FC**	**Costs without FC**	**Reduction**
1	20,703.69€	25,268.54€	18.07%
2	39,085.40€	44,675.95€	12.51%
3	34,756.43€	43,146.41€	19.45%
4	33,898.15€	39,255.22€	13.65%
5	34,947.55€	36,820.84€	5.09%
11	36,697.01€	38,082.93€	3.64%
12	29,618.69€	37,160.09€	20.29%

8.5% reduction in operational costs.

This translates to a reduction in the total energy used for cooling, decreasing from 23.22% to 19.33% of the data center's overall energy consumption. This corresponds to a 3.98% reduction in total energy consumption. In terms of environmental impact, emissions are reduced from 399.55 tonnes of CO_2 to 380.18 tonnes of CO_2.

The significant benefits of this implementation come as no surprise, as the return on investment is recovered in a maximum of three months when applied during the right months. This is why data center operators have widely adopted this approach.

3.2 Study Case 2

As this study case integrates both micro and macro scenarios, we establish a reference point to assess its potential benefits using the following situation: Data Center B does not have access to free cooling (operating under Micro Scenario 1), while Data Center D does have access to free cooling (operating under Micro Scenario 2). In this reference case, the two data centers are not interconnected via cloud services, meaning no load shifting occurs, and no macro scenarios are applied.

All further considerations for comparison will be made by evaluating Data Centers B and D as a whole. This means that all energy consumption and cost calculations will reflect the combined operation of both data centers.

The total operational cost for Data Centers B and D combined amounts to

Data Centre	**Total Energy [GWh]**	**Energy Used for Cooling [GWh]**	**Emissions [t_{CO_2}]**
B	1.571	0.3669	267.071
D	1.492	0.288	380.187
B+D	3.063	0.655	647.258

Table 3.3: Reference Case values

697,973.38.€. Next, we will examine the figures provided by the simulations for Study Case 2 to assess the impact of this scenario.

The values presented in Table **??** are also provided for Study Case 2, as shown in Table **??**. The reduction in energy consumption accounts for 0.09%

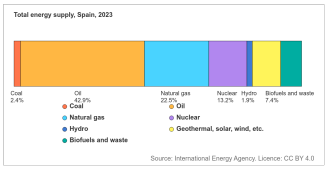
Table 3.4: Study Case 2 values

Data Centre	**Total Energy [GWh]**	**Energy Used for Cooling [GWh]**	**Emissions [t_{CO_2}]**
B	1.419	0.317	241.26
D	1.641	0.335	418.144
B+D	3.0602	0.652	659.404

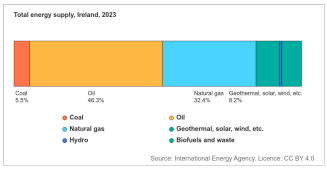
of the total energy consumption of the overall system and a 0.44% (reduction in cooling load consumption. Specifically, this corresponds to a decrease of 2.866 MWh in total energy consumption and 2.870.72 MWh in cooling energy consumption. As expected, these figures are nearly identical, given that the sole advantage of this configuration is the free cooling available in Dublin.

On the other hand, emissions are higher in Study Case 2 compared to our reference case, despite the lower energy consumption. The increase amounts to 1.88%, which is attributed to the energy mix differences between the two countries where the data centers are located. Specifically, the emission

factor for energy consumption is 0.2548 for Dublin and 0.17 for Barcelona.
newline This discrepancy is explained by the varying energy generation profiles of the two nations. Data Center B, located in Spain, benefits from a higher share of cleaner energy sources, including nuclear and renewables, which account for 18.96% of final energy consumption[20]. In contrast, Ireland has a lower reliance on cleaner sources, contributing only 12.69% to its final energy consumption.[21] As stated previously, 6.86% of the cooling load is shifted



(a) Spain's Energy Mix



(b) Ireland's Energy Mix

from B to D. Since Approach A1 represents a more conservative strategy within an already cautious load management assessment, the results align with expectations.
This initial analysis of the load management scenario suggests that shifting this portion of the load does not provide substantial benefits. Another key observation is that Data Center D accessed free cooling for a total of 425 hours, with the months benefiting from free cooling remaining consistent with Study Case 1.

The total cooling demand met by free cooling is 0.288 GWh, representing a 2.8% increase compared to the 0.281 GWh in the reference case. This small increase is due to D handling a higher cooling load while still operating under its usual cooling setup.

Furthermore, the operational costs amount to 708,777.04€, reflecting an increase of 1.55% compared to the reference case. All the data analyzed up to this point indicate that this setup (Study Case 2) does not offer any advantages over the initially considered reference case.

The limited gains can be attributed to the simulation methodology, which does not evaluate hybrid free cooling operations. Consequently, whenever free cooling is partially insufficient, the system fully switches to traditional cooling using the compressor unit, rather than dynamically blending both methods.

In a real-world application, however, the cooling tower pre-cools the water before it enters the chiller, allowing the chiller to operate at a reduced load. This enables a combination of partial mechanical cooling and partial free cooling, improving overall efficiency.

3.3 Study Case 3

Now, we will analyze the impact of Load Management Approach A2, which increases the load shifted from B to D by 2.95% of the total cooling load (B + D). This represents a shift from 6.86% in Approach A1 to 9.81% in Approach A2, effectively saturating D more than before. At first glance, we can observe

Table 3.5: Study Case 3 values

Data Centre	Total Energy [GWh]	Energy Used for Cooling [GWh]	Emissions [t_{CO_2}]
B	1.354	0.295	230.149
D	1.717	0.368	437.591
B+D	3.0712	0.663	667.740

that increasing the load, when compared to the numbers in the reference case and Study Case 2, does not result in any improvements in energy consumption, environmental and economical impact.

As seen in Table 3.6, all the listed values show increases rather than reductions. This further reinforces the previously mentioned limitation: increasing the cooling load in the simulation does not yield the expected benefits, as the model does not account for hybrid free cooling operations. A further

Value	Reference case	Study Case 2
Total Energy [GWh]	+0.26%	+0.36%
Energy Used for Cooling [GWh]	+1.24%	+1.68%
Emissions [t_{CO_2}]	+3.16%	+1.26%
Operational Costs	+5.80%	+4.18%

Table 3.6: Comparison of values between Study Case 3 and Study Case 2, as well as between Study Case 3 and the Reference Case.

confirmation of this trend can be observed in the reduction of hours during which D utilized free cooling, dropping to a total of 358 hours. The cooling demand satisfied by free cooling also decreased to 0.2473 GWh, marking a 14.13% decrease from Study Case 2 and an 11.99% decrease from the Reference Case.

All considerations made in Study Case 2 are perfectly applicable and bear the results seen.

3.4 Study Case 4

This final study case, which could even be termed the extreme case, places both B and D under Macro Scenario 2 and Micro Scenario 2 simultaneously.

Data Centre	Total Energy [GWh]	Energy Used for Cooling [GWh]	Emissions [Tco2]
B	1.329	0.270	225.889
D	1.717	0.368	437.592
B+D	3.046	0.638	663.480

Table 3.7: Values of Study Case 4

Chapter 4

Conclusions and Future Developments

The advantages of free cooling are undeniable, as demonstrated in Study Case 1, and this applies universally to any installation, regardless of location.

However, as observed across all study cases, under the current conditions, implementing load management between two data centers—despite differences in free cooling potential—does not necessarily yield advantages. In some instances, it may even worsen overall system performance. This is primarily due to two key factors: disparities in grid emission factors and electricity prices between the two locations. These elements significantly impact the feasibility and benefits of such an approach.

With the transition toward greener energy grids, the outlook may change. Many European nations are advancing toward net-zero carbon emissions, with Ireland, for example, aiming for 80% of its electricity to come from renewable sources in the near future. Additionally, nuclear energy is gaining traction, as seen in Microsoft's acquisition of Three Mile Island for potential data center energy supply.

Looking ahead, these shifts may open new avenues for optimization—such as leveraging waste heat from servers and heat pumps to generate additional energy. This integration could redefine the efficiency landscape of data center operations, offering sustainable and cost-effective solutions.

A hybrid architecture that integrates cloud, fog, and edge computing—combined with on-site green energy generation in regions with high free cooling potential—emerges as the most efficient configuration for optimizing the energy-intensive data industry.

By strategically deploying edge and fog servers—which have lower energy demands compared to cloud data centers—in locations without access to free cooling, the overall energy consumption of the geo-distributed data center network can be significantly reduced. This approach ensures that high-power workloads are handled in regions where energy efficiency is maximized, while lower-power computing is distributed to sites where free cooling is unavailable, minimizing energy waste and enhancing sustainability

Bibliography

[1] Otto Van Geet and David Sickinger. Best practices guide for energy-efficient data center design. Technical report, National Renewable Energy Laboratory (NREL), Golden, CO (United States), 2024.

[2] Andrea Mott Paficic Northwest Nationa Laboratory. Best practices for air-side economizers operation and maintenance, https://www.pnnl.gov/projects/om-best-practices/air-side-economizers, 2021.

[3] Andrea Mott Paficic Northwest Nationa Laboratory. Consider water-side economizers, https://www.energystar.gov/products/data_center_equipment/16-more-ways-cut-energy-waste-data-center/consider-water-side-economizers, 2021.

[4] Robert E McFarlane. Ashrae standards and practices for data centers. *Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center*, pages 175–191, 2021.

[5] Rodrigo S Couto, Stefano Secci, Miguel Elias M Campista, and Luís Henrique MK Costa. Latency versus survivability in geo-distributed data center design. In *2014 IEEE Global Communications Conference*, pages 1102–1107. IEEE, 2014.

[6] Aldo Canova, Giuseppe Laudicina, Paolo Lazzeroni, Nicolas Perez-Mora, and Maurizio Repetto. Exploitation and optimal management of free cooling for industrial refrigeration. *Energy Conversion and Management*, 198:111815, 2019.

[7] Jinkyun Cho and Yundeok Kim. Improving energy efficiency of dedicated cooling system and its contribution towards meeting an energy-optimized data center. *Applied Energy*, 165:967–982, 2016.

[8] IEA. Data centres and data transmission networks, https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks, 2023.

[9] Robert Arno, Addam Friedl, Peter Gross, and Robert J Schuerger. Reliability of data centers by tier classification. *IEEE Transactions on Industry Applications*, 48(2):777–783, 2011.

[10] Kuei-Peng Lee and Hsiang-Lun Chen. Analysis of energy saving potential of air-side free cooling for data centers in worldwide climate zones. *Energy and buildings*, 64:103–112, 2013.

[11] Alex Von Hassler DataSpan. Crac vs. crah cooling units: What's the difference?, https://dataspan.com/blog/crac-vs-crah-cooling-units-whats-the-difference/: :text=the

[12] Niru Kumari, Amip Shah, Cullen Bash, Yuan Chen, Zhenhua Liu, Zhikui Wang, Tahir Cader, Matt Slaby, Darren Cepulis, Carlos Felix, et al. Optimizing data center energy efficiency via ambient-aware it workload scheduling. In *13th InterSociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, pages 539–544. IEEE, 2012.

[13] Bilin Shao, Dan Song, Genqing Bian, and Yu Zhao. Network-aware data placement strategy in storage cluster system. *Mathematical Problems in Engineering*, 2020(1):5970583, 2020.

[14] Barzan A Yosuf, Amal A Alahmadi, Taisir EH El-Gorashi, and Jaafar MH Elmirghani. Cloud fog architectures in 6g networks. In *6G Mobile Wireless Networks*, pages 285–326. Springer, 2021.

[15] Jae-Gil Lee and Minseo Kang. Geospatial big data: challenges and opportunities. *Big Data Research*, 2(2):74–81, 2015.

[16] Raphael Eidenbenz and Thomas Locher. Task allocation for distributed stream processing. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9. IEEE, 2016.

[17] Majid Raeis, Ali Tizghadam, and Alberto Leon-Garcia. Queue-learning: A reinforcement learning approach for providing quality of service. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 461–468, 2021.

[18] Marzieh Malekimajd, Ali Movaghar, and Seyedmahyar Hosseinimotlagh. Minimizing latency in geo-distributed clouds. *The Journal of Supercomputing*, 71:4423–4445, 2015.

[19] Dave Sterlace ABB. How data centers can minimize their energy use, https://new.abb.com/news/detail/66580/how-data-centers-can-minimize-their-energy-use, 2023.

[20] IEA. Spain, https://www.iea.org/countries/spain/renewablesr, 2023.

[21] IEA. Ireland, https://www.iea.org/countries/ireland/renewables, 2023.