Stochastic Assessment of Industrial Loads in the Automotive Industry for Transformer Sizing: Contribution and Responsibility

Xabier Martín García

Madrid, Spain

Abstract

This study applies the Monte Carlo simulation method to support decision-making in the design of industrial electrical infrastructure under uncertainty. The case focuses on determining the optimal sizing of the main power transformer in an automotive manufacturing plant. Given the complexity of coordinating multiple power-consuming processes, each with different levels of demand, simultaneity, and uncertainty, the Monte Carlo simulation provides a probabilistic framework to estimate expected power requirements and their confidence intervals. By iteratively simulating consumption scenarios derived from statistical distributions, the method enables a more robust sizing of such technically and economically critical electrical equipment. Simulation results in values with enough confidence and robustness suitable for educated decisions in the sizing process. Also, two archetypes are used to categorize the involvement degree and nature of each load in the system.
Keywords: Monte Carlo simulation; industrial power systems; transformer sizing; substation transformer; decarbonization; uncertainty modeling; probabilistic design; electrical load forecasting; automotive manufacturing; decision support; energy optimization; electrification; contribution; responsibility.

1. Introduction

Industrial facilities, especially in the automotive sector, operate within highly dynamic and uncertain environments. Power demand is influenced by multiple interacting factors such as production schedules, maintenance operations, environmental conditions, and management decisions. Designing an adequate electrical supply system, therefore, requires not only deterministic data but also an understanding of probabilistic behavior.

Probabilistic methods have already proven valuable in other sectors, especially in renewable energy projects where weather-related variability creates similar challenges. This paper examines an automotive plant undergoing expansion and electrification as part of its decarbonization strategy. Unlike previous studies, this case incorporates additional variables specific to automotive manufacturing, demonstrating how uncertainty affects critical infrastructure decisions in this strategic industrial sector.

Traditional engineering design typically relies on conservative assumptions and fixed safety margins. While practical, this approach can result in **oversized** systems (driving up costs and energy losses) or **undersized** ones (risking production interruptions and equipment failures). Even experienced engineers working with decades of industry knowledge often make educated guesses when accounting for uncertainty. Probabilistic methods like Monte Carlo simulation offer a more deterministic way to model this uncertainty. Given that substation transformers cost around **1.5-2M €** each (with prices varying by up to 25% depending on cooling systems, specifications, site requirements, copper market prices, insulation class, and labor) and can represent nearly half the factory's electrical equipment budget.

Monte Carlo simulations use random sampling from statistical distributions to estimate the probability of different outcomes. For industrial power systems, this means generating multiple realistic demand scenarios by accounting for factors like equipment simultaneity, utilization rates, and load variations across production areas.

The objective of this paper is to present a robust method for optimal sizing of the transformer for the main substation by simulating thousands of possible combinations of electrical loads and operating conditions across several subsystems. This study focuses on a given system and its subsystems (building and contained loads), showing how integrating probabilistic models into the design process can improve decision reliability and reduce financial risk in major capital investments.

2. Context and motivations

This study is based on a real automotive manufacturing facility where multiple buildings and production lines run simultaneously, each with highly variable power demands. The plant is currently expanding while transitioning away from natural gas, replacing heating, drying, and HVAC co-generation systems with fully electric alternatives. At the same time, production lines are being modernized with increased robotization to meet newer industry standards. Together, these changes are dramatically increasing electrical load and introducing new layers of complexity to energy planning.

In practice, pinning down the plant's total power demand means coordinating estimates from multiple departments (production, maintenance, logistics, environment, health & safety, and factory management) each working with different priorities, data sources, and timelines. The production department, for example, might delay specifying which paint-drying technology they'll use until corporate approves the investment. Meanwhile, the environmental team imposes power caps based on emissions targets and efficiency regulations. Decision-making bottlenecks make things worse: Logistics won't commit to power estimates until Services finalizes the changeroom capacity, which itself depends on Logistics' staffing projections. These **circular dependencies** make it nearly impossible to establish a reliable power requirement through traditional deterministic methods.

This scenario is essentially a **multi-variable uncertainty optimization** problem. Each subsystem has a nominal power demand, an associated variability (σ), and a probability of being active at any moment (subject to logical constraints later to be discussed). The interaction between these probabilities creates an enormous number of possible operating states. Computing every combination explicitly would be impractical as the number of states grows exponentially with the number of loads (2^n).

This is where **Monte Carlo** simulation becomes valuable. By randomly sampling combinations of active and inactive loads based on their simultaneity factors, we can reproduce the stochastic nature of industrial energy consumption and generate working data for analysis. Each simulation represents one plausible operating scenario. Running thousands of these simulations produces a probability distribution of total power demand that reflects real-world variability.

This probabilistic approach enables to:

1) **Identify** the most likely range of transformer loading under normal operations.
2) **Quantify** confidence intervals for different design capacities.
3) **Evaluate** the risk associated with under- or over-sizing the transformer.

Lastly, the motivation behind this study is to demonstrate that Monte Carlo–based decision support allows industrial engineers to move beyond static load assumptions, incorporating **uncertainty** and system **interdependence** directly into the design of critical electrical infrastructure.

3. Methodology

The methodology is based on the application of the Monte Carlo simulation method to estimate the probability distribution of total power demand in an automotive manufacturing facility. The approach models each major energy consumer as a probabilistic variable characterized by its nominal load, uncertainty, simultaneity factor, and usage factor. The combined effect of all consumers is evaluated through iterative random sampling, generating a representative range of possible total demands for transformer sizing..

Each major electrical consumer within the plant—such as the body shop, paint shop, logistics, and assembly—is modeled as an independent random process characterized by several stochastic parameters. In this way, the simulation iteratively generates thousands of possible operating states, computes the total power demand for each, and derives statistical metrics from the resulting distribution for **high-level decision-making** purposes.

3.1. Data Sourcing and Characterization

The stochastic models require detailed characterization of each power-consuming subsystem in the factory. An example of a typical automotive industry factory:

A. Historical Measurement Data
   i. **Nominal power**: obtained from nameplate ratings and monitored average demand values.
   ii. **Power factor**: extracted from power quality measurements at sub-distribution panels.
   iii. **Usage factors**: computed as the ratio between average and nominal consumption over representative periods.
   iv. **Uncertainty**: defined from the observed standard deviation of historical power data; in well-controlled production lines, this is typically within ±1%.

B. Design and Engineering Specifications
   i. **Simultaneity factor**: derived from operational estimates. Simultaneity denotes the load on a specific conductor where the total downstream loads add up depending on their synchronous uptime. Total load in a distribution branch equals maximum simultaneity load, all at maximum workload, or factor 1.
   ii. **Power factor**: industries with higher mechanical power loads or electrical motors tend to have lower pf than those where there is less demand due to their nature.

C. Expert Elicitation and Departmental Input
   i. **Production efficiency**: operation schedules, shift patterns, technologies implemented at a given time.

D. Environmental and External Data
   i. **Ambient temperature**: used to model HVAC loads and Electrical Net thermal capacity.
   ii. **Relative humidity**: affects mostly HVAC calculations and drying processes.

In industrial environments, it is common to conceptually and physically discriminate **production**, **power**, and **lighting**, as these are usually segregated downstream for safety and homogeneity reasons. *Power*, in this context, refers to auxiliary power intakes typically destined for handheld machinery, computers, and other semi-permanent loads.

For instance, common Schuko sockets are rated for 16 A and 230 or 400 V AC, which means they can handle a maximum theoretical load (total installed power) of around 3 and 11 kW, respectively. However, this is not a realistic scenario. Firstly, protection devices are rated well below the maximum. Common residential electrical power contracts contemplate up to 3.7 kW (in Spain, at 230 V AC), but all aggregated power sockets will almost certainly never output 16 A at 230 V AC simultaneously. This principle also applies to industrial standards, with electrical socket boxes housing 2–8 sockets of 32 A each. The most common practice is to size protections slightly above 32 A, given that no special equipment will be connected. For this reason, **simultaneity and usage factors** are significantly lower for this kind of load.

Lighting technology dating from the 2000s is mostly based on electronic ballasts with high noise for indoor applications and high-intensity discharge lamps for outdoors. Current industry breakthroughs and environmental standards are displacing these technologies due to their high power load and impact on total energy demand. **LED** technology is now the most accepted and readily available solution, with a much higher light-to-load ratio. Besides, **smart building management systems** (BMS) increase efficiency through measurement and control software solutions. This upgrade can account for a decrease of up to **70% on average** in total lighting energy consumption.

Moreover, **distribution type** is also a key factor affecting the simulation. Most variables with low dispersion follow normal distributions, while others, because of their nature, work best with alternative types. However, for this case, only essential parameters will be used to keep it as clean and simple as possible, while making possible corrections and variations involving other factors.

TABLE 1. Parameters Affecting System Load Estimation

Parameter	**Symbol**	**Source**	**Distribution**	**Uncertainty**
Nominal power	P_i	Measured / datasheet	Constant	±1–5%
Simultaneity factor	$f_{sim,i}$	Expert input	Uniform / Beta	±10–20%
Usage factor	$f_{use,i}$	Measurement	Normal	±5%
Power factor	pf_i	Power quality data	Normal	±2%
Process efficiency	η_i	Design / estimation	Beta	±10%
Maintenance downtime	$t_{maint,i}$	Maintenance logs	Exponential	variable
Ambient temperature	T_i	Meteorological data	Normal	seasonal
Relative humidity	RM_i	Meteorological data	Normal	seasonal

These datasets were statistically analyzed to obtain mean values, standard deviations, and potential correlations for use in the Monte Carlo model.

3.2. Definition of Variables and Stochastic Parameters

The definition of the variables involved in the simulation should be an agreement between complexity, representation, realism and common sense. To start with, the following variables directly linked to electrical

The most clear and representative variables are:

- P_{inst}: nominal power (kW), installed machinery rating.
- f_{sim}: simultaneity factor[1].
- f_{use}: use factor.
- σ_i: standard deviation from nominal power.

More variables may be defined to make the model relatively more complex in order to represent real behavior of the load fluctuations. This includes parameters indirectly related to electrical phenomena such as those affecting HVAC and production demands, which translates into electrical load fluctuation. Some variables evolve with co-dependences rather than independently. It is important to assess which ones are correlated via first-hand datasets or operative experience directly from engineers and managers. However, due to reasonable lack of industrial datasets, some can be inferred by logical reasoning.

[1] The factor indicates the common load demand for protection dimensioning matters. A distribution bar that has 10 kW worth of equipment connected will seldom have every one of them active at the same time. Dimensioning with 0.3 factor makes it reflect the desgned rating (based on previous know-how and expectations of use) rather than the maximum or total installed power.

In this study, most common and representative correlations shall be considered as the purpose of the study is to apply an assessment methodology for cases with high uncertainty and lack of previous data.

3.3. Simulation Framework

For low level analysis, such as production lines, using distribution-based activation can be the most feasible approach. For each simulation iteration i, the activation status of each consumer is determined probabilistically according to its distribution function:

$$State_i = \begin{cases} 1, if\ random < F_{distr} \\ 0, otherwise \end{cases}$$

This formula first triggers the load depending on the previously estimated simultaneity factor. This is the most important aspect of Monte Carlo simulations, as it constitutes the main source of random combinations. However, this method has its flaws, since the combinations may have no logical correlation with real-world system behavior. For this purpose, further data refinement and scrutinizing are required to exclude illogical or impossible combinations (e.g., robots idle while soldering transformers are operating). Furthermore, large-scale systems are statistically almost always in an **on-duty state**, which makes this approach computationally less efficient and unnecessarily complicated.

The actual power consumed by each active load is modelled as a normalized random variables μ and δ:

$$P_{sim} = \sum_{i=0}^{n} P_i * \mu_i + \sum_{i=0}^{n} P_i * \delta_i$$

Here, both deterministic and stochastic factors are added. This allows each fixed load to fluctuate realistically within its uncertainty or variance range, capturing the stochastic nature of each power consumer.

Installed power can be estimated with usage factor to adjust actual power use, averaging based on empirical and know-how data:

$$P_i = P_{inst} * f_{use}$$

Power factor also plays a major role in uncertainty and adds its own randomness:

$$S_{sim} = \sum_{i=1}^{n} \frac{P_{sim}}{pf_i}, \qquad where\ pf_i = N(\mu_{pf}, \sigma_{pf}^2)$$

$$P_{max} = \sum_{i=1}^{n} Preal * fsim$$

Simultaneity can be implemented by performing ON/OFF state activations on different loads. However, this approach assumes that entire MW-consuming buildings will be either active (with a certain degree of variability) or idle, which is not realistic. The only way to address this issue is to apply the simultaneity factor as a downscaling factor, thereby losing fidelity to the real system.

For these reasons, **Gaussian Copula** method will be used instead, employing a symmetrical correlation matrix $R \in [-1,1]^{n \times n}$ for n total segregated loads. This matrix is built using expert knowledge on process/load correlations. For example, lightning and power outlet usage is highly and positively correlated due to workers using tools in lit areas most of the time.

Cholesky decomposition is used to calculate L such as:

$$R = LL^t \ where \ L \in \mathbb{R}^{n \times n}$$

To add the stochastic perturbations a vector Z is randomly generated before applying correlation:

$$\mathrm{Z} = Z_{i\,i=1}^{\;n} \ where \ \mathrm{Z} \sim \mathrm{N}(0,1)$$

Vector $X = LZ$ is operated so it now contains both randomness and correlation between variables. This also satisfies that $Cov(X) = R$.

Lastly, conversion to the previously estimated distribution is done as follows for the simulated P_i within a certain building using the probability integral transform. F_i^{-1} is the inverse cumulative distribution function (CDF) and maps X_i into the desired (previously expert-estimated) distribution, and is Φ the standard cumulative distribution function.

$$P_i = F_i^{-1}\big(\Phi(X_i)\big), i = 1{:}\,M$$

For each load, its nature shall be taken into account, again, based on expert criteria or historical data extracted from industrial datasets or observation:

$$\begin{cases} P_1 \sim Lognormal(\mu_1, \sigma_1^2) \\ P_2 \sim Beta(\alpha_2, \beta_2) \\ P_3 \sim \ldots \end{cases}$$

This approach is used for analyzing behavior inside specialized buildings. In other words, simultaneity applies between separate buildings is usually always 1 as they are all positively correlated because of the sequential nature of modern production in big factories. Some variables such as HVAC can have their randomness solidary for each iteration or be left out of the Gaussian Copula for later simulation. This reflects that some phenomena affect factory-wide loads, which is the case of the total light or HVAC needs. However, this has to be assessed for each case, for example, if a certain workshop generates enough heat to justify a significant increase or decrease in HVAC demand and diverts it via water or air transportation, though this is quite exotic in most cases.

The process is repeated for several thousands of iterations (typically in the magnitude of tens of thousands) to ensure enough definition and granularity in further statistical analysis methods. It is worth noting that maximum power demand is a statistically uncommon state to happen, while averages show more representative but unsafe indicators. So, a confidence interval (CI) will be used to indicate how probable is to find values within defined limits. For most cases, 95% is a safe CI to use in any not-so-critical industry, accounting for a decent criticality as the scope is the sizing of a transformer for an automotive factory, where most of the time there are only one or two levels of redundancy:

$$CI = \bar{x} \pm z \frac{\sigma}{\sqrt{n}} \rightarrow$$

3.4. Actual Scenario Definition

For this study, several general guidelines have been taken into account. The existing production facilities remain unchanged except for the gas oven decarbonization, which will consist of a mixture of technologies in old and new production chains. Compressed air, data, and water supply will be upscaled for increased capacity. This involves new IT, compressor, and pump loads from scratch due to demolition works on previous auxiliary facilities.

The three main expansions will consist of Body Shop, Paint Shop, and Final Assembly Line (hereafter abbreviated as **Body**, **Paint**, and **Final**). Prototyping exists as a separate unit for R&D purposes. Due to the nature of their processes, each has a different electrical load characterization:

1. **Body**: employs robots in magnitudes of hundreds and soldering transformers that demand peaks of very high current. Besides, conveyor and logistics related loads are relatively smaller due the initial production stage weight.
    a. Existing robots and soldering stations remain the same.
    b. Lighting usage is reduced due to unmanned operations in large areas of the building.
    c. Spot and MIG soldering stations will require cooling.
    d. Air extraction is specially required at this production stage for fumes and air renovation in closed spaces.
    e. Robots operate solder stations typically rated between 10-20 kVA.
    f. Hydraulic needs will account for a significant amount.
2. **Paint**: consists of a proprietary industrial solution made by a specialized company. Common technology for main gas oven replacement includes resistive electric heating, infrared curing and high efficiency heat pumps. However, the exact load characterization is largely unknown due to being a proprietary production process.
    a. Available information shows large demand of compressed air and process HVAC.
    b. Heat recovery is feasible and will be taken into account for a small reduction in demand.
    c. Lighting loads are slightly increased for production and quality requirements (imperfection detection).
    d. A mixture of several drying technologies will be used to replace former gas-powered processes. Experience shows that 60% is made of heat-pumps, 25% resistive heating and 15% of localized infrared heating. Heat recovery can save up to an estimated 20% once the process kickstarts, deducting from thermal inertial load heat-pumps deliver while cold.
    e. Compressed air and Volatile Organic Compound (VOC) treatment acts as a continuous load.
3. **Final**: overall less electrically demanding. It consists on a large logistic line where the vehicle has its most critical parts and components added, which means a significant increase in weight on each step.
    a. New forklifts will have bigger batteries, which in turn demand greater charging stations.
    b. Transportation and conveyor loads will be larger in comparison with other shops.
    c. Test-benches and quality assurance department are present at this stage, which increase electricity usage uncertainty even though they are relatively small loads.
4. **Prototyping**: separate building with Body and Final workshops. Their workload is way smaller and work in two shifts. They account for a moderate percentage of installed loads but with little usage factor, mostly HVAC consumer.

It is important to note that there are significant gaps in publicly available data. In real scenarios, this data is sourced directly from contractor companies and managed by the client's electrical department. Body Shop is perhaps easier to gain insight into due to being a process that involves mainly robots, soldering stations, and simpler conveyor routes. Final Assembly employs fewer robots, as the tasks require human manipulation and involve fewer risks, so realistic estimations are easier as it resembles a traditional manufacturing plant. Paint Shop process, however, is more dependent on external know-how, as the automotive industry focuses on vehicle design, not painting technology. These paint shop processes tend to be more complex, as many environmental conditions such as air quality, time, and paint type are critical and need to be assessed. Estimations in this area are not just optional but a key aspect of modeling factory power needs.

The statistical representation of industrial electricity consumption requires probability distributions that are both physically meaningful and flexible enough to reproduce real-world behavior. In this study, the behavior of electricity demand is modeled using a restricted set of continuous distributions, namely **Gamma**, **Lognormal**, and **Normal**. Distribution selection is based on the physical nature of the underlying processes and on empirical evidence.

Gamma distribution is adopted as the primary modelling choice for loads characterized by their additive behavior and relatively stable operation. When demand results from the superposition of many similar contributions, the Gamma distribution provides a natural representation, offering positive support and controllable skewness without imposing artificial upper bounds. This makes it particularly suitable for modelling base industrial loads with moderate variability.

Beta distribution applies in cases bounded by installation capacity. This choice ensures physical consistency, prevents artificial tail inflation, and reflects the regulated, state-driven nature of these loads, in contrast to event-driven processes better represented by Gamma or Lognormal distributions. Usually, these distributions are consistent and don't show peaks that contribute to load accumulation.

Lognormal distribution is used in loads whose demand is driven by multiplicative effects, transient peaks, or operating conditions that can amplify consumption non-linearly. Processes involving start-up phases, high-intensity cycles, or strong sensitivity to external factors tend to generate heavier right tails than those captured by the Gamma family. In such cases, the Lognormal distribution offers a more realistic representation of extreme demand values while keeping positivity.

Lastly, **normal** distribution is retained only for consumption components that show low variability and symmetric fluctuations around a well-defined means. In these cases, the probability of negative values is negligible compared to the mean due to its large size. Its use is therefore limited to near-constant loads or aggregated components with strong central tendency, such as lighting loads.

The following table summarizes chosen distributions and their characteristic parameters for the paintshop example:

TABLE 2. Loads and their Distribution, Load and Characteristic Parameters

			Normal		**Lognormal**		**Gamma**	
Load Types	**Distr.**	Mean load (kW)	μ	σ	k	σ	k	θ
Paintshop								
Ovens	Log	4875	0.75	20	8.47	0.20	25	195
Painting Booths	Beta	4000	0.50	15	8.28	0.15	44.44	90
Conveyors	Beta	375	0.50	5	5.93	0.05	400	0.94
Ventilation	Log	1125	0.75	10	7.02	0.10	100	11.25
Air Compressors	Gamma	810	0.45	15	6.69	0.15	44.44	18.23
Pumps & Filtration	Gamma	390	0.60	10	5.96	0.10	100	3.90
Robots	Beta	900	0.50	20	6.09	0.20	25	18
Spraying	Beta	1100	0.55	20	6.39	0.20	25	24.20
Heat Recovery	Beta	400	0.50	15	5.98	0.15	44.44	9
Pre-treatment	Log	650	0.40	20	5.54	0.20	25	10.40
HVAC	Normal	825	0.55	20	6.70	0.20	25	33
Lighting	Normal	450	0.90	5	6.11	0.05	400	1.13

Following the need for data to correlate these loads at the building level as mentioned before, a symmetrical correlation matrix R is elaborated from expert knowledge and empirical observations on loads. The fitness and validation of this data exceed the scope of this work, so values are chosen using common-sense and expert criteria for simulation purposes. Values exist within the range $[-1, 1]$ and indicate the strength and direction (positive/negative) of the correlation. Indices follow the same order as in the table above. For example, this matrix shows a high correlation of 0.8 between power outlets and lighting, as adequate lighting conditions are necessary for electrical equipment use.

$$
R_{paint} = \begin{pmatrix}
1 & 0.4 & 0.2 & 0.3 & 0.2 & 0.1 & 0.2 & 0.1 & 0.85 & 0.1 & 0.1 & 0 \\
0.4 & 1 & 0.3 & 0.75 & 0.5 & 0.5 & 0.5 & 0.6 & 0.3 & 0.3 & 0.1 & 0.05 \\
0.2 & 0.3 & 1 & 0.3 & 0.2 & 0.1 & 0.2 & 0.2 & 0.2 & 0.2 & 0.1 & 0.05 \\
0.3 & 0.75 & 0.3 & 1 & 0.4 & 0.4 & 0.4 & 0.5 & 0.3 & 0.2 & 0.1 & 0.05 \\
0.2 & 0.5 & 0.2 & 0.4 & 1 & 0.3 & 0.5 & 0.5 & 0.2 & 0.1 & 0.1 & 0.05 \\
0.1 & 0.5 & 0.1 & 0.4 & 0.3 & 1 & 0.3 & 0.5 & 0.1 & 0.1 & 0.05 & 0.05 \\
0.2 & 0.5 & 0.2 & 0.4 & 0.5 & 0.3 & 1 & 0.55 & 0.2 & 0.1 & 0.1 & 0.05 \\
0.1 & 0.6 & 0.2 & 0.5 & 0.5 & 0.5 & 0.55 & 1 & 0.1 & 0.1 & 0.1 & 0.05 \\
0.85 & 0.3 & 0.2 & 0.3 & 0.2 & 0.1 & 0.2 & 0.1 & 1 & 0.1 & 0.1 & 0 \\
0.1 & 0.3 & 0.2 & 0.2 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 1 & 0.05 & 0.05 \\
0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.05 & 0.1 & 0.1 & 0.1 & 0.05 & 1 & 0.2 \\
0 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0 & 0.05 & 0.2 & 1
\end{pmatrix}
$$

L matrix represents Cholesky decomposition, so $R = LL^T$ is satisfied, being L a lower symmetrical matrix of the same dimension as R. Be noted that this method does not allow many or large negative correlations due to its nature.

$$
L_{1paint} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.4 & 0.91 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.2 & 0.24 & 0.95 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.3 & 0.68 & 0.07 & 0.65 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.2 & 0.45 & 0.05 & 0.03 & 0.86 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.1 & 0.50 & -0.04 & 0.04 & 0.05 & 0.85 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.2 & 0.45 & 0.05 & 0.03 & 0.28 & 0.04 & 0.81 & 0 & 0 & 0 & 0 & 0 \\
0.1 & 0.61 & 0.03 & 0.07 & 0.22 & 0.19 & 0.21 & 0.68 & 0 & 0 & 0 & 0 \\
0.85 & -0.04 & 0.04 & 0.10 & 0.05 & 0.03 & 0.03 & 0.01 & 0.50 & 0 & 0 & 0 \\
0.1 & 0.28 & 0.11 & -0.05 & -0.06 & -0.04 & -0.04 & -0.07 & 0.06 & 0.93 & 0 & 0 \\
0.1 & 0.06 & 0.06 & 0.03 & 0.05 & 0.007 & 0.03 & 0.03 & 0.01 & 0.02 & 0.98 & 0 \\
0 & 0.05 & 0.03 & 0.01 & 0.02 & 0.026 & 0.01 & -0.00 & -0.01 & 0.03 & 0.19 & 0.97
\end{pmatrix}
$$

As mentioned in the framework section, next steps involve generating random Z_i vectors of same dimension as R with normal distribution $\sim N(0,1)$ as a product of Z_i and L. Then, U_i is calculated as normal distribution function $\Phi(X_i)$ to add randomness to th $F_i^{-1}(\Phi(X_i))$

Starting from a correlation matrix R, its Cholesky factor L is computed such that $R = LL^{\mathsf{T}}$. A vector of independent standard normal variables Z is generated and transformed into correlated normals $X = LZ$. Finally, the standard normal cumulative distribution function $N(0,1)$ is applied component-wise, resulting in a vector $U_i = \Phi(X_i)$ of uniformly distributed variables in that preserves the dependence structure defined by R.

However, it should be noted a key aspect when working with Cholesky decomposition and R correlation matrices. Even though it exceeds the scope of the article, **eigenvalues** quantify the independent variance directions of the correlation matrix. Near-zero or negative eigenvalues indicate over-constraints or inconsistency, which prevents a stable Cholesky factorization and causes to “fail” decomposition because negative correlation “extracts” variance and inserts cancellation, so as to say. Moreover, this is what prevents the Cholesky matrix to be **completely singular**.

Therefore, negative correlations were restricted to preserve strictly positive eigenvalues of the correlation matrix, ensuring the existence and numerical stability of the Gaussian copula used for dependency modelling.

3.5. Validation and Consistency Check

In the absence of empirical measurements, validation is limited to internal consistency checks, including recovery of theoretical moments, upper-quantile behavior, and dependence structure, which are sufficient for assessing the numerical correctness of the Monte Carlo framework.

Analysis of Monte Carlo outputs by consumption category highlights marked differences in variability and tail behavior. Categories with higher **coefficients of variation** and **percentile-to-mean ratios** (tail amplification factor) contribute to distorting improbable measurements at both ends, depending on their distribution. This reinforces the relevance of **category-level stochastic modeling** when assessing aggregate load uncertainty.

4. Results and Discussion

The proposed probabilistic framework was applied to generate synthetic electricity consumption data for the analyzed industrial processes from the electrical design engineer's point of view. The resulting simulated individual and aggregated loads show statistically coherent behavior at both the individual and aggregated levels, keeping internal correlation. As mentioned before, the focus will be on the Paintshop building for the sake of brevity.

General metrics for this building and their distributions are the following, all dimensions in kW except skewness and kurtosis metrics:

TABLE 3. Statistical Moments of Paintshop Loads

Load	Mean	STDev	Var.	Median	P95	Skew.	Kurt.
Ovens	3 727	3 781	0.28	4 834	7 669	0.65	0.51
Painting Booths	4 162	4 222	0.53	3 954	5 944	0.67	0.39
Conveyors	2 727	2 767	0.25	526	566	0.06	0.05
Ventilation	4 572	4 638	5.44	1 119	1 248	0.22	-0.09
Air Compressors	2 636	2 674	2.67	801	1 013	0.33	0.16
Pumps & Filtration	2 982	3 025	1.09	317	446	0.52	0.52
Robots	3 281	3 329	0.28	447	584	0.33	0.02
Spraying	3 875	3 875	0,62	596	754	0.32	-0.09
Heat Recovery	3 480	3 480	1.43	402	447	-0.10	0.42
Pre-treatment	3 551	3 551	0.92	258	3 42	0.49	0.27
HVAC	3 338	3 338	1.28	1031	1 373	-0.39	-0.51
Lighting	2 965	3 008	1.01	451	485	-0.03	0.01
Total	14 931	2 200	0.15	14 946	18 926	1.15	0.75

The aggregated electrical demand exhibits a clear right-skewed distribution, with a mean value of approximately **14.9 MW** and a P95 of **18.9 MW**, representing an increase of about 27% over the mean. This confirms the presence of non-Gaussian distribution and justifies the use of a probabilistic approach for system sizing.

At the subsystem level, several loads show comparable mean values but markedly different distributional characteristics, including variability, asymmetry, and tail behavior. While large consumers such as ovens, HVAC, and painting booths dominate average consumption, only ovens display statistical features consistent with a partial role in amplifying high-load scenarios. Most remaining loads exhibit low medians and limited P95 values, indicating intermittent operation with marginal impact on extreme aggregated demand. Overall, the results suggest that peak demand emerges from the combined behavior of multiple loads rather than from a single dominant driver.

Key findings include:

1. Total demand distribution is right-skewed, with a **P95 approximately 27% higher than the mean**.
2. Ventilation, HVAC, and painting booths contribute strongly to average load, but their moderate skewness and flat kurtosis indicate stabilizing, base-load-type behavior.
3. Ovens show the **highest P95-to-mean ratio** among major consumers, identifying them as partial drivers of high-load scenarios.
4. Several subsystems present low medians and limited P95 values, reflecting intermittent operation with negligible influence on system extremes.
5. Negative or near-zero kurtosis in regulated systems (HVAC, ventilation) suggests bounded operation and a low propensity to generate extreme events.

Compared to a non-simulated classical approach where values would be calculated from the product of installed power and simultaneity and usage factors, this method uses the **95th percentile (P95)**. This percentile captures high-load operating conditions that occur regularly while excluding very rare events that would lead to excessive overdimensioning. This is common practice in engineering for critical systems that allow some flexibility on risk without sacrificing realism.

In this case, utilization factors times installed power equals around **15 MW** in the paint shop. Further downgrading for simultaneity and "know-how" factor corrections distorts the final estimate. However, the calculated P95 is approximately **18 MW**, which represents a considerable difference of ~**20%**.

The next graph illustrates how the total simulated and aggregated loads for each iteration are distributed. Note the multimodality, tail weights and limit cases:

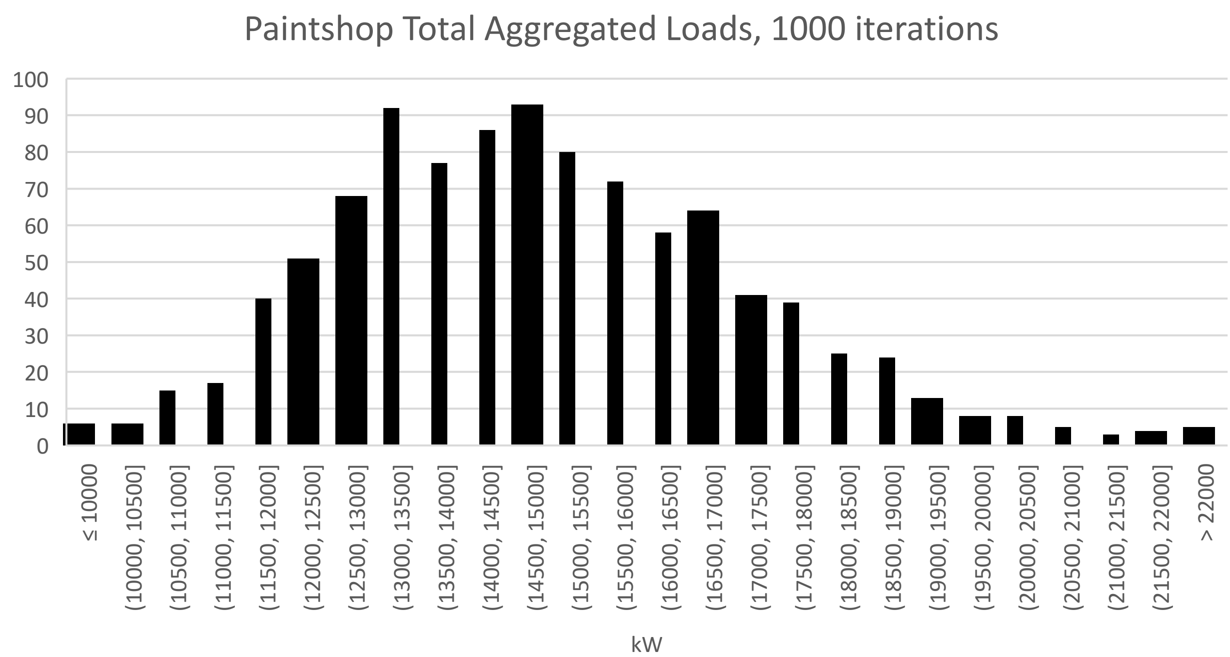

Figure 1. Paintshop Total Aggregated Loads, 1000 iterations

4.1. Marginal Distributions

Histograms for each marginal distribution are shown to better illustrate how this affects the aggregation of loads. Four have been selected to represent each distribution type and scale: **pumps**, **HVAC**, **ovens**, and **conveyors**.

Gamma distribution adequately represents loads constrained to positive values with right-skewed distributions such as pump loads. These are expected to work as quasi-continuous loads with very small variances in governing parameters such as flow, density, temperature and pressure.

Beta distribution is used in loads where switching off is not usually feasible and continuous modulation is used instead. HVAC loads can be characterized so due to its operation nature, a bounded and modulated type of load that are not driven by events affecting extremes. In this case, as the distribution function yields a 0–1 value, it is used to scale the installed total load plus a 10% minimum. This ensures that consistency is kept due to HVAC never operating at full or minimum load (except from exceptional cold switch-ons). Also, α and β are controlled to fulfill $\alpha + \beta < 6$ for stability and $(\alpha, \beta) > 1$ for concavity (unimodal type). With these parameters, the histograms present a platykurtic multimodality that is consistent with a left-skewed (due to minimum operating power) and evenly distributed operating conditions (due to external factors affecting load modulation).

Lognormal distribution captures higher dispersion and asymmetric behavior with larger tails. Key difference with Gamma-governed loads is the multiplicative and batch-operation nature, hence keeping the left asymmetry, but following more concentrated patterns. Oven loads are a good example of this due to its operation needing big batches of workload to be efficient.

Lastly, **normal** distributions were (and usually are) suitable for stabilized or aggregated consumptions with approximately symmetric behavior such as lighting, given no special conditions. In some cases, Gamma and Lognormal distributions may show little distribution differences against Normal distributions due to superposition and small variances among large samples (n >1000). Conveyors can easily fit in this category due to their simple nature due to having less variations in energy consumption.

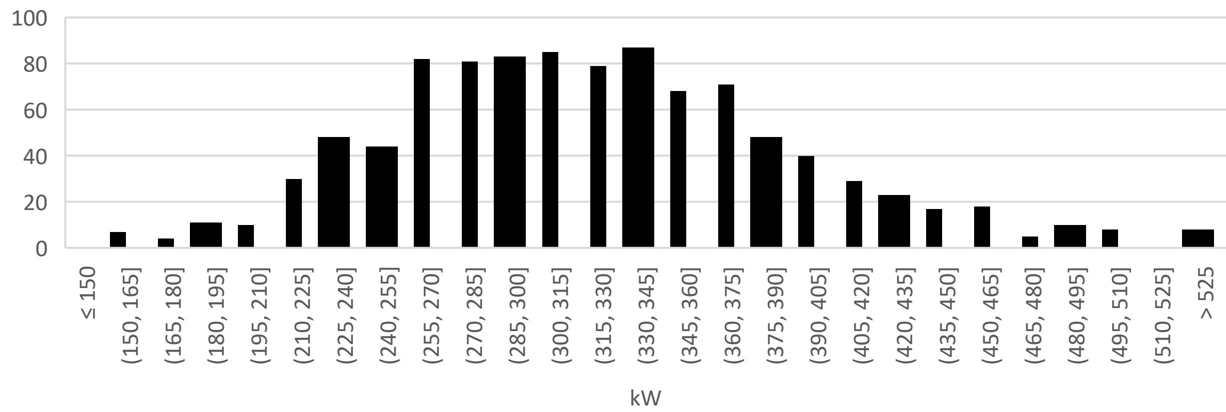

Figure 2. Pump Loads, Gamma Distribution

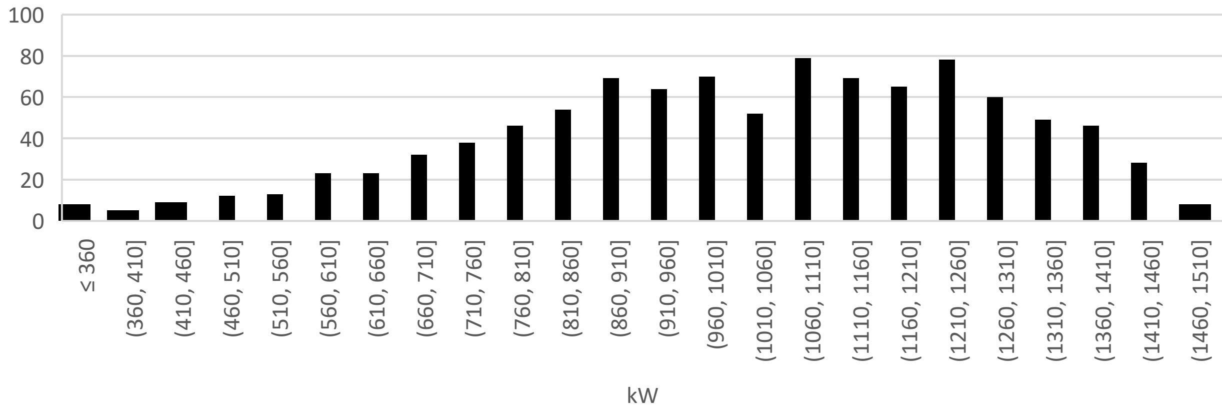

Figure 3. HVAC Loads, Beta Distribution

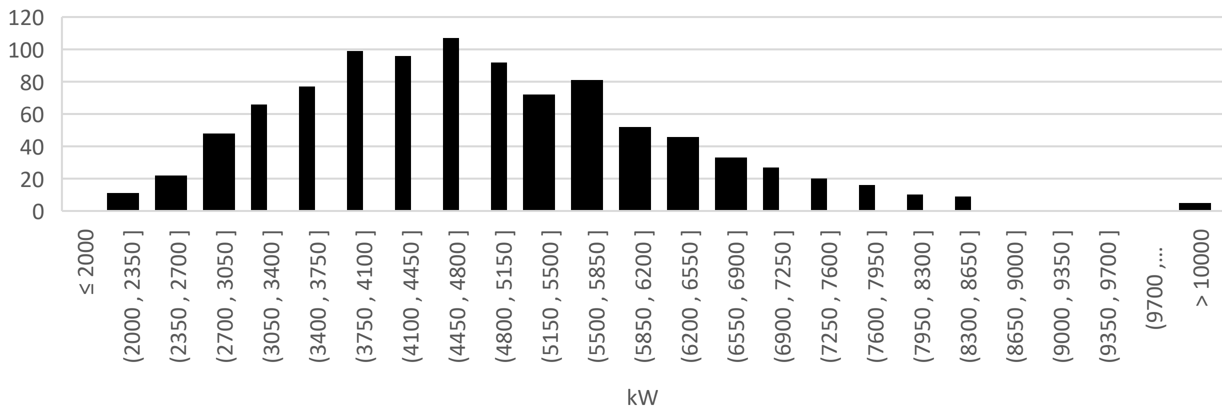

Figure 4. Oven Loads, Lognormal Distribution

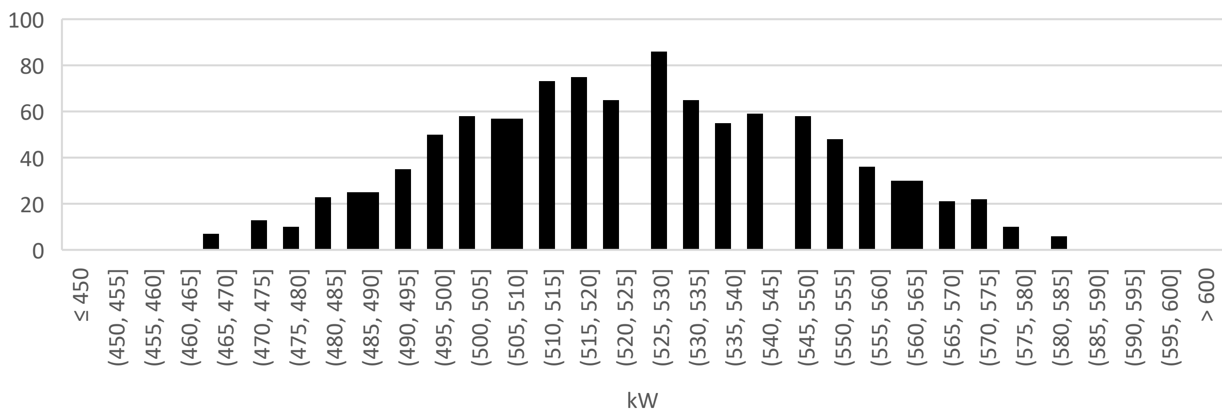

Figure 5. Conveyor Loads, Normal Distribution

Overall, the simulated consumption profiles present realistic variability, plausible inter-load relationships, and statistically consistent marginal behavior, supporting their use in downstream energy analyses. Note that the shown histograms display multimodal effects in their own distributions. This is expected to occur once all loads are aggregated due to superposition to obtain the final probability **spectrum**. However, due to Cholesky decomposition and the correlation of loads, the nature of these histograms has been altered.

Next, a scatter plot has been used to show the imposed correlation structure between significant load pairs. In this case, oven loads are the most representative ones due to their magnitude. Correlation can be seen between oven and total loads as a relatively straight line, representing the weight of the former rather than its obvious proportionality. Heat recovery loads show less but still identifiable correlation, as they grow when more heat is present in the process. Conveyors are mostly independent, as their roundness shows. Distortions in shape are consistent with the marginal transformations applied to Normal, Gamma, and Lognormal distributions and do not affect the preservation of the dependency structure. Note that total aggregated loads have been downscaled **40 times** for representation purposes.

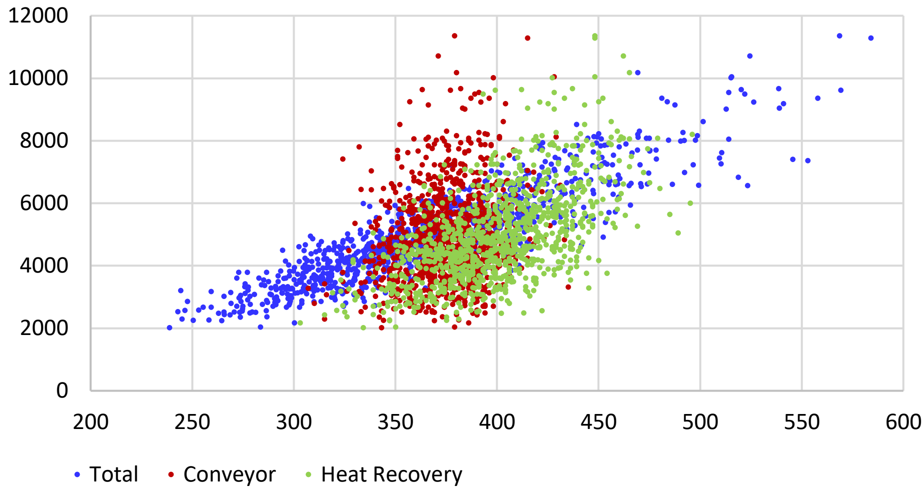

Figure 6. Oven Correlation to Loads

4.2. Contribution vs Responsibility

Final thoughts on this framework raise a question regarding **contribution** and **responsibility**. This may provoke confusion about observed contribution and inherent responsibility for load increase. The first metric refers purely to mean and median values, stable or base states. Responsibility is what drives peak consumption patterns, which may be classified accordingly. It is very important to note what loads are involved and how they participate in the total distribution when addressing contribution and responsibility.

For this purpose, the following histogram represents the most common load magnitudes. As expected in manufacturing industry, most building-related loads don't usually exceed 1 installed MW. A few categories fall beyond 3 MW, which account for oven and painting booth types of loads. There is also a notable plateau between 3 and 4.8 MW. These loads are most likely the ovens and paint booths mentioned before, with some extreme cases found from 6 MW onwards.

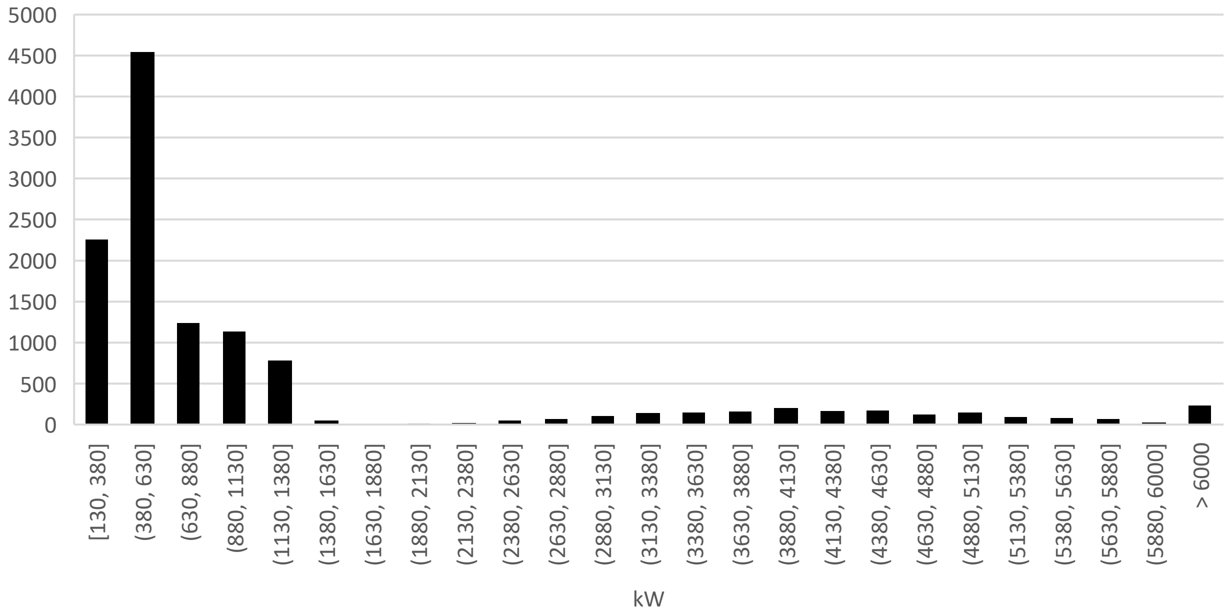

Figure 7. Paintshop Load Characterization by Bins

This gives a hint on how many smaller loads are present on the process, but judging only from this characterization no responsibilities can be extracted.

The following histograms of total electrical load, where bar height represents the frequency of each load bin. The colored lower segment indicates the average relative contribution of the asset conditioned on that total-load interval, rather than a share of the frequency.

4

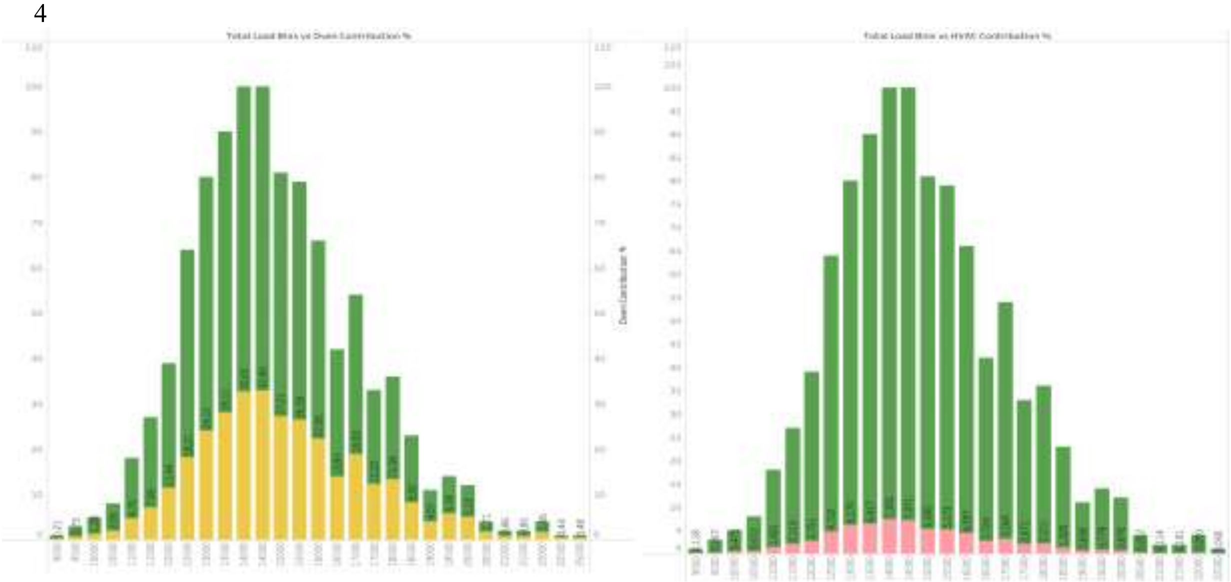

Figure 8. Relative HVAC & Oven Contribution % in Total Aggregated Load Histogram

This combined representation highlights different electrical roles. Baseline-defining loads, such as **ovens**, exhibit higher relative contribution around the central bins where the system operates most of the time, but a decreasing presence toward the upper tail, indicating limited influence on extreme demand events. In contrast, risk-driving loads, such as **HVAC**, maintain or increase their relative contribution in higher-load bins, revealing stronger association with peak conditions.

The divergence between total-load frequency and conditional contribution is therefore intrinsic to the conditioning approach and should be interpreted as evidence of distinct contribution and responsibility regimes, rather than as a visualization artifact.

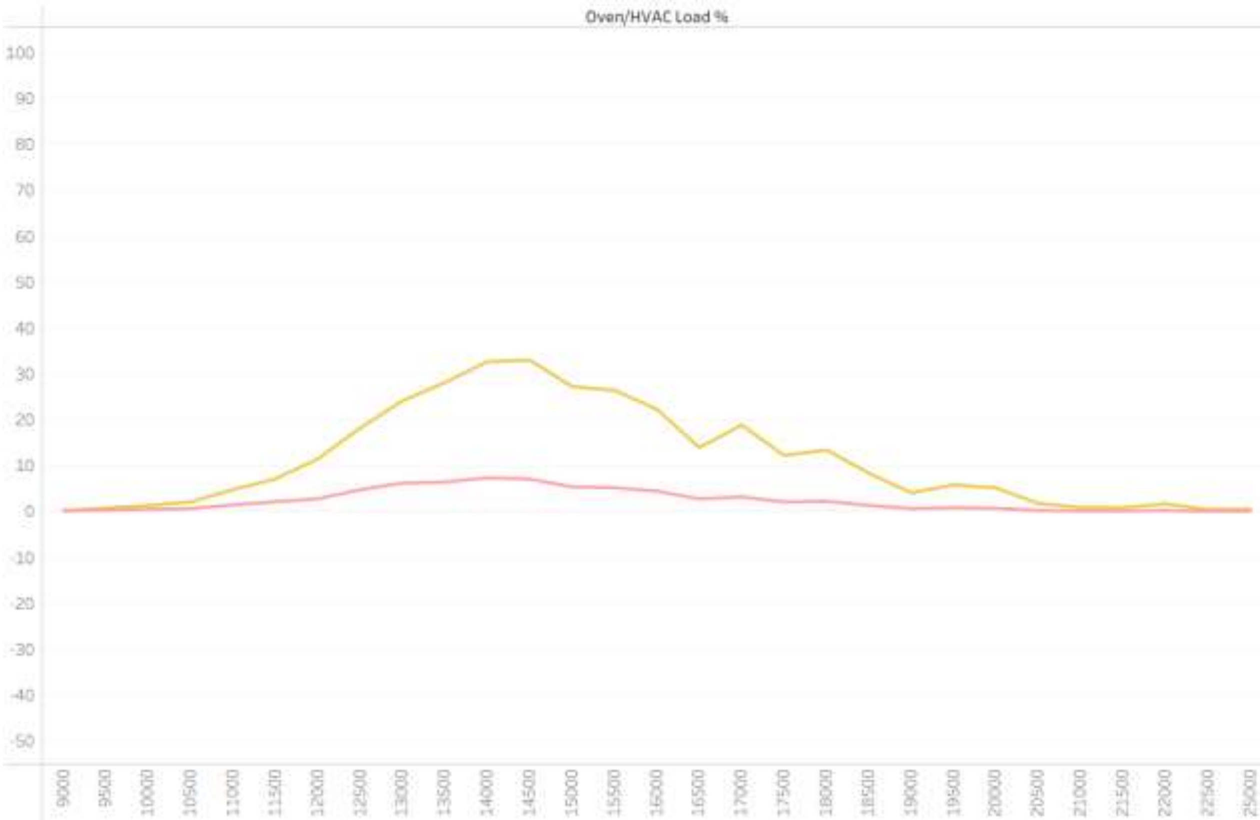

Figure 9. Oven to HVAC Load in %

Following this line, two formal metrics are proposed for contribution and responsibility based on observed data.

Load **contribution** is the fraction of a given load's 95th percentile over the total one. This is a confident way of expressing how much at 95% of the total cases the load contributes to establish a base consumption-line or base-line drivers.

$$C_i^{95} = \frac{P95_i}{P95_{total}}$$

Load **responsibility** is quantified relative to the system design percentile (P95 of the aggregated demand). For each load category, responsibility is defined as the relative reduction in P95 when the load is removed from the aggregation, thereby capturing its role as a driver of extreme demand scenarios:

$$R_i^{95} = \frac{P95_{total} - P95_{excluding\ i}}{P95_{total}}$$

TABLE 4. Contribution and Responsibility Metrics

Parameter	Ovens	HVAC	Conveyor	Pumps
C_i^{95}	40.6%	7.4%	3%	2.3%
R_i^{95}	36.2%	6.1%	3%	2.1%
R_i^{95}/C_i^{95}	0,89	0,82	1	0,90

This study case shows that $\boldsymbol{C} \propto \boldsymbol{R}$. This means that the model is statistically stable and healthy. Different results can be expected when working with different distributions and regimes, such as exceptional momentaneous loads (magnetizers and soldering stations) or long-lasting controlled processes (such as chemical ones). However, in this case, there is a significant difference the higher the contribution levels, as can be seen in ovens, **4%** difference between contribution and responsibility.

For this load, there can be said that greatly helps to establish a consumption baseline, more than what it adds to peak and extreme case values. That is, counter-intuitively, ovens help to distribute risk, not concentrate it, HVAC even a bit more so. Only conveyors may pose a threat, but their installed power is small in the scale and well balanced (based on initial assumptions). Extreme demand scenarios are therefore driven by the collective behavior of the system rather than by a single dominant load.

In this study, **HVAC** shows relatively **high responsibility** compared to its contribution, reflecting its tendency to be active or ramped during high-demand conditions. However, even HVAC remains proportional in this dataset, indicating a lack of dominant peak-driving behavior.

Findings imply that electrical loads shall be categorized in the two following groups to better assess their nature and involvement in system sizing analyses:

1. High **contribution** assets:
    a. Consistent across time.
    b. Dominate the median and lower quantiles of the distribution.
    c. Define the baseline and typical operating condition.
    d. Shift the entire load duration curve when removed.
2. High **responsibility** assets:
    a. Over-represent in high-percentile load conditions.
    b. Correlate strongly with total-load ramps or peaks.
    c. Reshape the upper tail of the load distribution when conditioned.
    d. Drive demand charges and infrastructure constraints.

However, when two parameters are combined, two-dimensional regimes appear. This is an adequate enough metric to stablish the degree of involvement of the load in the system once iterated This case occupies Regime I (Baseline Definers) for **ovens** and Regime IV (Electrically Marginal) for **conveyors and pumps**, while **HVAC** approaches Regime II without fully decoupling responsibility from contribution.

The following table summarizes the four archetypal electrical regimes that emerge when contribution and responsibility are analyzed separately:

TABLE 5. Four Proposed Regime Quadrants

Regime	Contr.	Resp.	Distribution	Behavior
I. Baseline Definers	High	Low	Dominant in mid/low Qs	Stable load, long operation time
II. Risk Drivers	Low	High	Sparse in central, over in >P95	Intermittent, correlated with peaks/ramps
III. System Dominators	High	High	Dominant overall, upper tail	Variable but large loads tied to production
IV. Electrically Marginal	Low	Low	Limited presence	Small, balanced, weakly coupled loads

This way, the proposed metrics are robust enough to assess implications in the sizing of a transformer at estimation or bidding phases to gain statistical confidence on the topic. This approach also helps with communication between chiefs of the different electricity-consuming departments to better assess loads defining a clear objective. That is, completing the contribution/responsibility chart for sizing purposes.

5. Conclusion

This study presented a probabilistic approach for transformer sizing in industrial facilities through Monte Carlo simulation. By replacing deterministic load assumptions with stochastic representations of simultaneity, usage, and environmental variability, the proposed method captures the inherent uncertainty of real industrial operations.

5.1. Data Sourcing

The case study results show that total demand follows a clear probabilistic pattern, with distinct clusters corresponding to different production scenarios. This allowed to define transformer capacity not as a single fixed value, but as a confidence-based design target. In the real project, **60 MW** was used as the upgrade target value in place of the old 30 MW transformer. This approach provides a statistically sound basis for design decisions, balancing the risks of over- or under-sizing while accounting for real operational variability. For the paintshop case, as most of the paintshop process is proprietary and difficult to initially estimate, this was a good way for kick-start estimations that helped medium-voltage cables and low-voltage board sizing calculus.

The methodology also effectively handles correlated variables through covariance-based sampling. This is a substantial improvement over traditional safety-factor methods, which can't explicitly capture these interdependencies, only by engineering-coefficient[2] application.

Ultimately, this probabilistic framework demonstrates that Monte Carlo-based load assessment can be a practical and robust tool for industrial energy planning. It enables more rational infrastructure sizing, better cost optimization, and greater adaptability for future changes, advantages that experience-based approaches alone simply can't deliver with the same level of confidence.

[2] Be it understood as coefficients in the 1.2-1.5 range meant to add arbitrary but know-how-driven safety margins.

5.2. Managerial Implications

From an operational perspective, this probabilistic framework gives plant designers and energy managers a more structured way to make decisions. Rather than relying exclusively on conservative safety margins, they can now size transformers and infrastructure based on **quantified risk levels** derived directly from statistical demand distributions.

This approach optimizes investment by matching transformer capacity to realistic operating probabilities, avoiding unnecessary oversizing costs while maintaining reliability. For scenarios where uncertainty remains high, specific critical combinations can be tested to validate the logic behind design choices.

The methodology also improves communication between technical teams and management. Results expressed in probabilistic terms, like "95% confidence of staying below capacity", are easier to understand and defend than abstract safety factors. This supports evidence-based planning and helps build more credible long-term electrification strategies.

6. Future Works and Further Investigation

This work opens up several interesting directions for future research. The approach could be applied to other combinatorial problems where randomness creates too much complexity for traditional deterministic methods. And because it's computationally inexpensive once you've identified and documented your variables properly, it could be useful across different industries and scientific fields. Here are some possibilities that branch off from this study:

1. Variable classification, validated distribution. Given there are yet unexplored fields, variables need to be accordingly validated to assure used distributions are fitting for the behavior they represent.
2. Applications in key sector workflows and production (pharma, aero, agriculture).
3. Usage in fields where empiric data is scarce and probabilistic solutions are needed (history): for tasks than involve working with unreliable numbers and/or little data, as cited in some antiquity records describing events, methods or army compositions.
4. Data acquisition mechanisms: with sufficient historical SCADA data, machine-learning regression or Bayesian inference could be used to estimate distribution parameters and correlations automatically, reducing the reliance on expert elicitation.
5. Inclusion of new technologies in the model: renewables, energy storage systems, intelligent or smart systems (such as cities, grids, transportation means) add up yet to come challenges with their own uncertainty in terms on power consumption, deployment and integration.
6. Integration with economics and budget metrics: Extending the model to include CAPEX/OPEX trade-offs and equipment reliability would support multi-objective optimization of transformer sizing. Coupling stochastic demand with Weibull-modeled lifetimes could capture both economic and aging effects in asset planning, besides failures and other physical risks.

7. References

1. **Albuquerque, D. M., & Cavalcanti, G. D. (2022).** Stochastic analysis of distribution transformer loading: A Monte Carlo approach. *Energies*, 15(4), 1321.
2. **Bracale, A., Carpinelli, G., & Proto, D. (2018).** A probabilistic approach for capacity assessment of electrical assets in industrial networks. *Applied Sciences*, 8(12), 2635.
3. **Jordehi, A. R. (2018).** How to deal with uncertainties in electric power system problems: A review. *Renewable and Sustainable Energy Reviews*, 96, 145-155.
4. **Aien, M., Hajebrahimi, A., & Fotuhi-Firuzabad, M. (2016).** A comprehensive review on uncertainty modeling techniques in power system studies. *Renewable and Sustainable Energy Reviews*, 57, 1077-1089.
5. **Molinaro, V., & Spazzafumo, G. (2020).** A Monte Carlo Approach for the Sizing of Power Systems with Energy Storage. *Energies*, 13(15), 3901.
6. **Vahidinasab, V., Mohammadi-Ivatloo, B., Mehdizadeh, M., & Ghazvini, M. A. (2020).** Uncertainty modeling in power systems: A review of methods and applications. *Renewable and Sustainable Energy Reviews*, 111, 110-125.
7. **Xie, S., Hu, Z., & Zhou, D. (2020).** Multi-objective optimization for industrial power system design under demand uncertainty. *Energies*, 13(12), 3121.
8. **Li, Z., Zhao, T., & Zhang, L. (2019).** Probabilistic sizing of distribution transformers considering the impact of low-carbon technologies. *IEEE Access*, 7, 126985-126994.
9. **Tascikaraoglu, A., & Sanandaji, B. M. (2016).** Probabilistic load forecasting in distribution networks: A review of Monte Carlo methods. *Journal of Modern Power Systems and Clean Energy*, 4(4), 543-556.
10. **Pfeiffer, P., & Dalmer, A. (2022).** Developing industrial load archetypes for energy system modeling and decision-making. *Energies*, 15(9), 3218.