VALIDATION OF MSCI CARBON EMISSIONS DATA LEVERAGING WEPP FOR ELECTRIC POWER GENERATION COMPANIES

by

Fu Zheng

Submitted to the Department Management, Technology, and Economics in partial fulfillment of graduation requirements for the degree of

Master of Science

ETH Zurich

September 2021

Supervisor: Didier Sornette

Signature: _____

Co-supervisor: Vahid Moosavi

Signature: _____

Abstract

There is a rising importance of carbon emissions for policy setting, business, and research, but some concerns about data quality have appeared, including uncertainty, inconsistency, verification, and compliance. As one of the most influential data providers of carbon emissions, Morgan Stanley Capital International (MSCI) is frequently used by a wide range of stakeholders, such as governments, investors, and researchers. However, the quality of MSCI's carbon emissions data is not validated. This thesis investigates the MSCI data quality and the potential reasons for the data quality issues leveraging the scope 1 carbon emissions dataset established based on the World Electric Power Plants (WEPP) dataset by using emission factors and the production model, and the Carbon Disclosure Project (CDP) dataset. Through correlation analysis and statistical tests (Spearman's rank correlation test, Wilcoxon signed-rank test, and Kruskal-Wallis H test), the results reveal that there is a monotonic moderate correlation existing between the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI, but there is a significant difference between the distributions of these two samples. As for the consistency between MSCI and CDP, scope 1, scope 2, and scope 1 and 2 of both carbon emissions and carbon intensity are highly consistent, demonstrating that the information companies disclose to the public is consistent with the information they disclose through CDP. But scope 1 and scope 1 and 2 have a higher consistency than scope 2. As expected, the consistency is higher for the companies that MSCI directly cites from CDP or estimates based on CDP than those both in MSCI and CDP datasets. However, MSCI and CDP have different distributions for carbon emissions and carbon intensity in all scopes. The data discrepancies may stem from raw data sources, methodology (e.g., capacity factors, energy sources classification, and the

ownership structure), and the company matching quality between different datasets. Based on the results, we cast doubt on the MSCI data quality at least for relatively pure electric power generation companies. Our findings suggest that all the stakeholders should pay attention to the data quality issues and validate the data through alternative independent data sources before applying it. To improve the data quality in carbon emissions and facilitate the carbon reduction process, greater transparency in data collection and reporting, and comparable data sources are needed.

CONTENTS

CHAPTER I. INTRODUCTION

1.1 Problem statements and challenges

Anthropogenic activities have contributed notably to the negative consequences of climate change by increasing the concentration of greenhouse gases (GHGs) in the atmosphere [1].

Over the past few years, carbon footprint has grown in prominence in the financial sector as a technique of measuring and disclosing carbon emissions (greenhouse gas emissions) from both internal operations and investment portfolios. Carbon emissions and carbon intensity are utilized to project possible future scenarios, as well as the future changes in population, economic activities, and energy technologies, such as those used in the Intergovernmental Panel on Climate Change (IPCC) assessments.

Corporations themselves report the majority of carbon emissions and carbon intensity data, normally without independent third parties' validation on the accuracy of this information. The data quality must be evaluated to provide high-quality data for transparent GHG monitoring and reliable future scenario prediction.

1.2 Aims and objectives

As one of the main GHG data providers, the carbon emissions and carbon intensity provided by Morgan Stanley Capital International (MSCI) are widely used by international organizations, researchers, institutes, and investors. Despite the wide usage of MSCI carbon emissions data and the recognized importance of the data on the policy setting, research, and business, the MSCI data quality is usually not validated before applying it. Hence, in this thesis, we mainly use the electric power plants data, provided by World Electric Power Plants (WEPP), to establish the scope 1 carbon emissions dataset to validate the corresponding data in the MSCI dataset for

electric power generation companies. In the meantime, we leverage the Carbon Disclosure Project (CDP) data to assess the data consistency between MSCI and CDP.

1.3 Thesis structure

This thesis is divided into six sections. Following this introduction, Chapter II gives some background information on the data quality issues that exist in carbon emissions and carbon intensity data, as well as their potential influence on research, business, and policymaking. The four data sources used in this thesis are introduced in Chapter III. Continuing in Chapter IV, we introduce the methodology to validate the MSCI data quality through constructing a scope 1 carbon emissions dataset based on the WEPP data. The validation results are reported in Chapter V, along with a detailed explanation based on data analysis. In Chapter VI, we explore the potential causes of the results. Conclusions based on our work and a vision for future research directions are presented in the final chapter.

CHAPTER II. LITERATURE REVIEW

2.1 What are the carbon emissions?

Carbon emissions are the total GHG emissions created directly and indirectly by an individual, event, organization, service, place or product, and are measured as carbon dioxide equivalent (CO_2e) using the corresponding 100-year global warming potential (GWP100) [2].

Carbon emissions are also expressed as carbon footprint, GHG emissions, carbon equivalent emissions, or carbon dioxide equivalent emissions. In this thesis, we have primarily used carbon emissions to denote GHG emissions.

Per GHG Protocol, carbon emissions are grouped into three categories known as scope 1, scope 2 and, scope 3 carbon emissions. Scope 1 carbon emissions are direct emissions from owned or controlled sources. Scope 2 carbon emissions are indirect

emissions from the generation of purchased energy. Scope 3 carbon emissions are all indirect emissions (not included in scope 2) that occur in the value chain of the reporting company, including both upstream and downstream emissions [3].

Scope 1 carbon emissions are direct emissions occurring from sources that the institution owns or controls, including on-campus stationary combustion of fossil fuels, mobile combustion of fossil fuels by institution-owned or controlled vehicles, and fugitive emissions. Fugitive emissions result from intentional or unintentional releases of GHGs, including the leakage of hydrofluorocarbons (HFCs) from refrigeration and air conditioning equipment as well as the release of methane (CH_4) from institution-owned farm animals. Scope 2 carbon emissions are indirect emissions from the purchased energy generation, i.e., indirect emissions generated in the production of electricity consumed by the institution. Scope 3 carbon emissions encompass all indirect emissions (not included in scope 2) that occur in the value chain of the reporting company, including both upstream and downstream emissions, i.e., any other indirect emissions that are a result of the institution's activities but occur from sources that the institution does not own or control, such as commuting; waste disposal; embodied emissions from extraction, production, and transportation of purchased goods; outsourced activities; contractor-owned vehicles; and line loss from electricity transmission and distribution [3, 4].

2.2 Importance of carbon emissions

Since 1850, each of the last four decades has been successively warmer than the decade before it. Human-caused emissions drive the observed warming, with greenhouse gas warming partially concealed by aerosol cooling. Human activity has indisputably warmed the climate, oceans, and land. There have been widespread and rapid changes in the atmosphere, ocean, cryosphere, and biosphere. As the near-linear

relationship between cumulative carbon emissions and the increase in global surface temperature shows, every tonne of carbon emissions contributes to global warming. The stabilization of carbon-induced global surface temperature increases requires achieving worldwide net-zero carbon emissions, with anthropogenic carbon emissions balanced by anthropogenic carbon removals [1].

To address the detrimental effects of climate change, 196 countries agreed in December 2015 to reduce global warming to well below 2 degrees Celsius, preferably 1.5 degrees Celsius, compared to pre-industrial (1861-1880) levels through the Paris Agreement. The Paris Agreement requires all member countries to lower their carbon emissions and to strengthen their efforts in the coming years [5].

There is an increasing importance of carbon emissions for policy setting, business, and research. Cities and organizations' carbon footprints have been quantified and reduced through legislative action, which is playing an essential role in policymaking [6]. Aside from legislative issues, the carbon footprint has become extremely important for business since it is linked to financial activities and the business world has predicted a carbon-constrained economy shortly [6].

2.3 Carbon emissions data quality problems

There are a rising number of voluntary or regulatory-driven initiatives, which require companies to disclose their GHG inventory following the GHG Protocol [7].

Despite the generalized use of a common protocol, there is an increasing concern with the figures provided by corporations. The carbon emissions data is published or self-reported by the companies, instead of evaluating by independent bodies, thus it’s extremely self-serving, biased, and unreliable. With the increasing availability of data and information on carbon footprint, some concerns about data quality have arisen, including uncertainty, inconsistency due to variability of parameters, models, and

approaches, and verification and compliance under both proposed and legislated schemes aimed at reducing human-induced global climate impact [6, 8].

2.3.1 Data uncertainty and inconsistency

Uncertainty about the carbon footprint data reported is an important parameter. Large uncertainties hinder progress in implementing, monitoring and verifying effective mitigation strategies [9].

The methodology used in different datasets, as well as the reported carbon emissions data, varies significantly.

There are three major concerns related to the quality of self-reported carbon emissions data: (1) the self-reporting bias introduced by voluntary investor-friendly reporting, limited data availability, (2) diversity of carbon emissions measurement methods and the disclosure recommendations and standards, raising the risk of companies intentionally greenwash their business activities, (3) inconsistency of reported data across different data providers [10].

In the absence of mandatory reporting, around half of companies in the coverage disclose their carbon emissions voluntarily, and thus estimation of carbon emissions data is needed for the remaining half of companies, leading to ineffective investor actions whose purpose was to moderate climate change [11]. Assessing the accuracy of the estimated carbon emissions data, investors are at least 2.4 times less likely to identify the worst 5% of emitting companies when using estimated carbon emissions data as compared to using self-reported data [10].

Important assumption disparities may not always be effectively addressed by researchers and policymakers due to the complexity of detecting and correcting gaps among various agency techniques and assumptions for energy and carbon statistics [12]. These distinctions are frequently overlooked, and presuming that the data given

is fungible may lead to erroneous comparisons and inconsistent conclusions across different datasets. If uncertainties and discrepancies are not effectively taken into account, these unmentioned uncertainties have the potential to undermine policy aims and scientific study outcomes.

According to the study by Macknick [12], if not completely understood, data disparities in energy statistics and CO_2 inventories can have a significant impact on climate modeling inputs as well as national and international policies that rely on precise estimates of carbon emissions. The study also found that depending on which datasets and methodologies are used to calculate emissions, intranational and international carbon emissions trading programs, such as the US Regional GHG Initiative and the EU Emissions Trading System, could have significantly different allocations of carbon emissions, and thus financial outcomes for individual member parties. Besides, the study revealed that while the uncertainty surrounding emissions from fuel combustion has been well documented in many cases, the inclusion of nonfuel combustion emissions, such as emissions from cement production and land-use change, to the assessments of total anthropogenic impacts on the carbon cycle and carbon policy decisions, could add to the uncertainty. If new national carbon taxes were implemented in specific countries or credits were apportioned to carbon-emitting activities, much consideration would have to be given to what emissions, both fuel combustion-related and nonfuel combustion-related emissions, would be taxed or given credits, as well as how to collect, monitor and assess the data uncertainty [12].

2.3.2 Data verification and compliance

The common resources used in carbon footprint calculations are standards of greenhouse gas accounting, while footprint verification is not required [6].

As for the criterion “verification” when assessing how well the carbon-related datasets match the users’ expectations, needs, and preferences, stakeholders from intergovernmental organizations were most skeptical and only 6.7% would agree to use data without third-party verification while stakeholders from companies (private and state-owned) preferred to use data with third-party verification, with 58.3% agreeing [9].

Moreover, the ability of the international community to provide some independent verification of emissions inventories, or at the very least the ability to falsify some components of an emissions inventory, would considerably boost confidence in an international treaty [13].

2.4 Potential data quality problems in MSCI

MSCI is one of the biggest carbon emissions data providers. However, the quality of the carbon emissions data provided by MSCI is not validated, and the uncertainty inhabited in MSCI carbon emissions data is not assessed.

Besides, one of MSCI's main data sources is CDP and CDP also has a certain degree of uncertainty. Several studies have raised concerns about the quality of data collected and published by the CDP, regardless of its success [11]. As stated in the study by Faria [8], uncertainty estimates on total gross direct emissions (scope 1 carbon emissions) are one of the CDP dataset's issues. The study has shown that less than 16% of companies in the CDP dataset have reported uncertainty figures higher than 10%, which is an acceptable uncertainty threshold used in the European Trading Scheme, and 8% of companies reported other things such as “no uncertainty” and “unknown”.

2.5 Carbon emissions in the energy sector

Due to IPCC negotiations, GHG emission inventories have recently become more important in domestic and international policy settings, and many states have enacted policies to reduce national energy usage and GHG emissions, as this sector is typically the largest contributor of GHG emissions and statistics from this sector are readily available with a low level of uncertainty [7].

As reported in the study by Ge Friedrich and Vigna [14], energy consumption is by far the most significant source of human-caused GHG emissions, accounting for 76% (37.2 $GtCO_2e$) of global emissions. The energy sector includes transportation, electricity and heat, buildings, manufacturing and construction, fugitive emissions, and other fuel combustion. Within the energy sector, the largest source of emissions is the heat and electricity generation (15.6 $GtCO_2e$ in 2018, or 31.9% of total GHG emissions), followed by transportation (6.9 $GtCO_2e$ in 2018, or 14.2% of total emissions), and manufacturing and construction (6.2 $GtCO_2e$, or 12.6% of total emissions) [14]. Focusing on CO_2 would allow independent checks (with a lower than 10% uncertainty) on fossil-fuel combustion and deforestation, which account for three-quarters of GHG emissions under the United Nations Framework Convention on Climate Change (UNFCCC) [9].

CHAPTER III. DATA SOURCES

We derived a list of the datasets that are used in the following analysis, including MSCI, WEPP, CDP, and Orbis. Table 1 shows the carbon emissions or intensity type, reporting time, total company coverage, the source of the data, and the data granularity of each data source. The information is obtained through the data providers' websites, the methodology, and the datasets they provided.

Table 1: Overview of the data sources, i.e., MSCI, WEPP, CDP, and Orbis, including the different types of carbon emissions and carbon intensity in the dataset, the reporting time (calendar year) of the carbon-related information, the number of companies in the dataset coverage, the original information source used by the data provider, and the data granularity.

Data source	Carbon emissions & carbon intensity type	Reporting time	Company coverage	Source	Data granularity
MSCI	Scope 1 carbon emissions, scope 1 and 2 carbon emissions, scope 1 and 2 carbon intensity	2018, 2021	13 681 (2018), 12 055 (2021)	Corporate sources, CDP, government databases, MSCI estimation	Company-level
WEPP	NA	2018	33 281	Primary and secondary database sources	Company-, plant-, unit -level
CDP	Scope 1 carbon emissions, Scope 2 carbon emissions, scope 1 and 2 carbon intensity	2021	7158	CDP annual questionnaire	Company-level
Orbis	NA	NA	Around 29 million companies	NA	Company-level

3.1 MSCI

MSCI ESG Research collects carbon emissions data for the companies in its coverage universe. Data is collected once per year from most recent corporate sources, including Annual Reports (AR), Corporate Social Responsibility Reports (SR), or websites. In addition, MSCI ESG Research uses the carbon emissions data reported through CDP or government databases when reported data is not available through direct corporate disclosure. When companies do not disclose data, MSCI ESG Research uses proprietary methodologies to estimate Scope 1, Scope 2, Upstream Scope 3, and Downstream Scope 3 carbon emissions [4].

In this thesis, we use two MSCI datasets, one is from 2018 and the other is from 2021. In the following analysis, we use MSCI (2018) to indicate the MSCI dataset from 2018, and MSCI (2021) to indicate the one from 2021. All the year indicated in this thesis is the calendar year instead of the fiscal year.

For MSCI (2018) dataset, the number of unique companies is 13 681. While for MSCI (2021) dataset, the number of unique companies is 12 055. Both of the datasets include the scope 1 carbon emissions, scope 1 and 2 carbon emissions, and scope 1 and 2 carbon intensity data.

3.2 WEPP

Among all industries, the energy industry is by far the largest source of anthropogenic GHG emissions, with electricity and heat generation accounting for the majority of GHG emissions [14]. Electricity and heat production (31.9% of 2018 total GHG emissions [14]), i.e., the burning of coal, natural gas, and oil for electricity and heat, is the largest single source of global GHG emissions [15]. Fossil fuels account for about half of the electricity generated by thermal power plants. Fossil fuels (coal, oil, and gas) are still the predominant energy source for electricity generation [16].

The WEPP is selected owing to its frequent usage in academic and policy studies of the energy sector, and its coverage of power plants, in terms of capacities and fuel types. WEPP is one of the most widespread power plant databases used by academics, NGOs, and businesses and it is greater in terms of the number of units and represented capacities, as well as geographical scope, amount of detail, and definitions (e.g., fuel types), compared to other commonly used energy databases, i.e. Carbon Monitoring for Action (CARMA), European Network of Transmission System Operators for Electricity (ENTSOE), DOE Energy Storage Exchange (only pumped storages)

(ESE), Global Energy Observatory (GEO), Open Power System Data (Conventional Power Plants) (OPSD), World Resources Institute (WRI) [17].

The S&P Global Market Intelligence World Electric Power Plants Data Base (WEPP) is a comprehensive, global inventory of electric power generating units. It contains ownership, location, and engineering design data for power plants of all sizes and technologies operated by regulated utilities, private power companies, and industrial or commercial auto-producers in every country and major territory in the world, including units that are currently installed, projected, retired, or canceled. The WEPP is maintained and reissued quarterly in its entirety by S&P Global Market Intelligence, part of S&P Global Inc [18].

Direct surveys, vendor reference lists, power company financial and statistical reports, and websites, and the trade and business press are all sources of data for power plants. Primary sources are preferred, such as surveys and information created directly by owners, operators, and suppliers. Power plant data is retrieved, crosschecked, entered, and confirmed to the degree possible from various primary and secondary database sources.

Electric power plant data are obtained from numerous sources, including direct surveys, vendor reference lists, power company financial and statistical reports and web pages, and the trade and business press. Power plant data are retrieved, crosschecked, entered, and verified to the degree possible from various primary and secondary data sources, and the primary sources such as surveys and materials directly produced by owners, operators, and suppliers are used preferentially [18].

Information in the WEPP database is included at the company, plant, and unit levels. Units belong to plants, and plants belong to companies. One plant can have many electric power units, and one company can have many power plants. Company

data include the company name, electric type, and business type. Plant (site) location data include the city, state or province, country, geographic area, subregion, and postal code. Unit data include unit name, operating status, capacity (MWe), year-on-line, primary and alternate fuels, equipment vendors for the boiler (or reactor), turbine and/ or engine, and generator/ alternator, steam conditions, pollution control equipment, engineering and construction contractors, and cooling system data [18].

The WEPP dataset used in this analysis is from 2018 and contains information about 33 281 parent companies at 111 969 power plants with 220 478 units, among which 37 001 units are in commercial operation.

3.3 CDP

The CDP is a not-for-profit charity that runs the global disclosure system for investors, companies, cities, states, and regions to manage their environmental impacts [19]. Its goal is to encourage investors, companies, cities, states, and regions to share more information about their climate-change-related risks and opportunities, thereby making environmental reporting and risk management a business norm and driving disclosure, insight, and action toward a more sustainable economy [20].

The CDP holds the largest database of primary corporate climate change information in the world [8]. It is the largest climate change-focused data collection and assessment program, requesting information on GHG emissions, energy use, and the risks and opportunities from climate change from the world's largest companies in various capital markets via an annual questionnaire [21, 22]. Currently, there are companies, cities, states, and regions from over 90 countries disclosing through CDP on an annual basis [19]. Since 2002, over 8400 companies have publicly disclosed environmental information through CDP [20].

The CDP dataset used for the analysis is from 2021, including 7158 companies.

3.4 Orbis

The Orbis is Bureau van Dijk's flagship company database. Bureau van Dijk, a Moody's Analytics firm, is a significant business information publisher, and a specialist of private company data combined with software for searching and analyzing companies [23].

The Orbis dataset contains information on companies across the world, around 400 million companies, with focuses on private companies as well as presenting companies in comparable formats. Bureau van Dijk gathers the information from over 170 different providers and claims that it brings extra value by standardizing the data and connecting the sources [23].

The Orbis dataset is used by CorpIndex [24], mainly using fuzzy name matching algorithm implemented by Swiss Re, and contains data about roughly 29 million companies. The Orbis dataset includes the company's ID, company name, country, revenue, workforce, industry, etc. It's in form of multiple JSONs which are collected by data scraping.

CHAPTER IV. METHODOLOGY

For the whole analysis covered in this paper, we only focus on scope 1, scope 2, and scope 1 and 2 carbon emissions and carbon intensity due to the data availability and precision. Most companies only disclose scope 1 and scope 2 carbon emissions. Around 475 world's largest companies disclosed their carbon footprints through CDP (2009), but around 83% of those participating companies only reported scope 1 and 2 carbon footprint, while the scope 3 carbon emissions of 5.8×10^9 tonsCO_2e were much higher than combined emissions of scope 1 and scope 2 (0.6 and 3.6×10^9 tonsCO_2e, respectively) [6]. In addition, the MSCI (2018), MSCI (2021), and CDP datasets, provided by MSCI and CDP, used for this analysis only contain scope 1 and

scope 2 carbon emissions. Even though the scope 3 emissions left the largest footprint, the data accuracy was the lowest [6].

4.1 WEPP-based scope 1 carbon emissions dataset construction

GHG data could be gathered in two ways: directly on-site real-time measurements or estimates based on emission variables and models. The best technique to use is determined by the goal (mandatory, voluntary, or for internal management), credibility, feasibility, cost, and capacity factors [6].

It is uncommon to quantify GHG emissions directly by measuring concentration and flow rate. Emissions are frequently estimated using a mass balance or stoichiometric basis unique to a facility or process. The use of documented emission factors and models are, nevertheless, the most popular and widely used approaches for estimating GHG emissions [6, 25]. In many situations, reliable emission statistics can be derived from fuel consumption data, especially when direct monitoring is unavailable or prohibitively expensive [25].

The WEPP dataset provides detailed electric power plants information, including company name, unit name, operating status, fuel type, capacity (MWe), etc. To leverage the WEPP data to validate the quality of the MSCI scope 1 carbon emissions data, we use the following five steps to construct the WEPP-based scope 1 carbon emissions dataset leveraging emission factors. The method to establish the dataset can be broken down into five steps: (1) translation of the terms in the WEPP dataset, (2) mapping of the fuel types in WEPP to higher-level fuel types, (3) calculation of the full load hours for each fuel type, (4) aggregation of individual electric power plant units into power plants and parent companies, (5) estimation of scope 1 carbon emissions.

Step 1. Translate the terms in the WEPP raw database

The first step aims to translate the abbreviation of key columns in the WEPP dataset, i.e., current unit status and fuel types, into full names based on the abbreviation list WEPP provided [18].

Step 2. Map the fuel types in WEPP to higher-level fuel types

The 62 fuel types in the WEPP dataset need to be mapped to one of the encompassing energy source or higher level of fuel types groups: biofuel, coal, geothermal, H_2, Helium, hydro, liquid fuel, natural gas, nuclear, solar, waste, or wind. The details about the mapping and classification of each fuel type in WEPP to energy source groups adopted are shown in Table 2.

Step 3. Calculate the full load hours for each fuel type

The capacity factors differ significantly depending on the plant and fuel type [26], and it's unrealistic to get the exact capacity factors of each fuel type for each plant in the WEPP dataset. The average capacity factor can be determined for any form of electricity-generating installations, such as a fossil fuel consuming power plant or one that uses renewable energy, and it can be used to compare different types of electricity generation [27]. Hence, we use the average capacity factor of each fuel type for all the electric power plants in the WEPP dataset, shown in Table 3. The full load hours are calculated based on the capacity factor, given by

$$Full\ load\ hours\ per\ year\ (h) = Capacity\ factor \times 365 \times 24\ h/year. \quad (1)$$

Step 4. Aggregate individual electric power plant units into power plants and parent companies

Each data record in the WEPP dataset represents the information of an individual electric power plant unit. The listed company in WEPP is both the facility operator and sole or majority owner, while the parent company in WEPP is used to track multinational power companies and holding companies and may also contain listings

for two or more companies in the joint venture or other arrangements [18]. However, MSCI reports individual companies, i.e., the parent companies stated in the WEPP dataset, so an aggregation of individual units into power plants and parent companies in the WEPP dataset is needed to ensure the comparability with the companies in the MSCI coverage universe. At the same time, the installed capacity per fuel type for each company is also aggregated to get the fuel capacity matrix for each company.

Besides, the WEPP dataset includes electric power units in different operation statuses, but only the units that are currently in commercial operation status contribute to the existing carbon emissions, thus we only focus on these electric power units for the whole analysis.

Step 5. Estimate scope 1 carbon emissions based on the WEPP dataset

According to the definition of different carbon emissions scopes, the estimated carbon emissions based on the WEPP dataset for electric power generation companies are scope 1 carbon emissions. Scope 1 carbon emissions based on WEPP for each company could be estimated using

$$Carbon\ emissions = \sum_i Emission\ factor_i \times Full\ load\ hours_i \times Power\ generation_i, \quad (2)$$

where i indicates the fuel type i in a sample company power generation mix.

The full load hours per year and the emission factor of each fuel type are listed in Table 3.

Table 2: Mapping and classification of the fuel types in the WEPP dataset to the energy source, i.e., the higher level of fuel types groups.

Energy source	Fuel types in the WEPP dataset
Biofuel	Wood or wood-waste fuel, refuse (unprocessed municipal solid waste), bagasse, biomass excluding wood chips but including agricultural waste and energy crops, pulping liquor (black liquor), landfill gas, biogas (produced by anaerobic digestion of biodegradable materials in closed systems), bioderived liquid fuels such as palm oil or vegetable oils or biodiesel or bio-oil or other bioliquids, sewage digester gas, syngas from gasified refuse, syngas from gasified wood or biomass, paper mill waste or sludges or wastepaper, waste paper and/or waste plastic, ethanol, wastewater sludge, methanol, syngas from gasified agricultural waste or poultry litter, meat and bonemeal, lignin (a wood polymer)
Coal	Coal, coal syngas (fuel for IGCC plants from gasified coal), petroleum coke, oil shale, peat, synthetic gas from petroleum coke, bitumen or asphalt or asphaltite, coal-water mixture (aka coal-water slurry), Orimulsion (trade name for emulsified bitumen manufactured in Venezuela - production terminated in 2006)
Geothermal	Geothermal
H_2	Hydrogen gas
Helium	Helium
Hydro	Water
Liquid fuel	Fuel oil, gasified crude oil or refinery bottoms or bitumen, kerosene (also see jet fuel), naphtha, liquified petroleum gas (usually butane or propane), jet fuel (typically kerosene, also naphtha-type), dimethyl ether, tar sands
Natural gas	Natural gas, liquified natural gas, blast-furnace gas also converter gas or LDG or finex gas (approx. 10% of the heat content of pipeline gas), refinery off-gas, flare gas or wellhead gas or associated gas, coal seam gas (aka coal bed gas or coal bed methane or CBM), mine gas (methane from active or abandoned coal mines), coke oven gas (approximately 50% of the heat content of pipeline natural gas), waste gas or low calorific gas from refineries or other industrial processes, corex process offgas, natural gas liquids (also natural gas condensate), top gas
Nuclear	Uranium
Solar	Solar power
Waste	Waste heat, scrap tires, industrial waste or refinery waste, hazardous waste, medical waste, manure fuel
Wind	Wind-powered turbines

Table 3: The capacity factors [28, 29, 30, 31, 32, 33, 34, 35, 36, 37], full load hours per year calculated using Eq. (1), and the scope 1 emission factors [4] of each fuel type or energy sources after the mapping and classification of the fuel types in the WEPP dataset to the higher level of fuel types groups in step 2.

Fuel Type	Capacity factor	Full load hours (h/year)	Scope 1 Emission factor (tonsCO_2e/MWh)
Coal	40%	3504	1.02
Liquid Fuel	30%	2628	0.758
Natural Gas	30%	2628	0.515
Hydro	45%	3942	0
Solar	18%	1577	0
Wind	40%	3504	0
Biofuel	70%	6132	0
Nuclear	81%	7096	0
Geothermal	85%	7446	0
H_2	50%	4380	0
Waste	35%	3066	0

4.2 MSCI (2018) validation leveraging WEPP

4.2.1 WEPP and MSCI (2018) datasets matching

Because there is no common column in the original WEPP and MSCI (2018) datasets, we link the separate WEPP and MSCI (2018) dataset together for the same company through the CorpIndex [24], mainly using fuzzy name matching algorithm implemented by Swiss Re, to ensure the companies in different datasets can be compared to each other. In the meanwhile, additional features are added from Orbis to benefit further analysis.

The matching process between two different datasets denote as A and B, through CorpIndex [24] mainly has the following 3 steps: (1) matching and enrichment dataset A with CorpIndex [24], (2) matching and enrichment dataset B with CorpIndex [24], (3) matching dataset A and B through the common and unique values introduced by CorpIndex [24].

We enrich and match the companies in WEPP and MSCI (2018) to companies in Orbis separately through CorpIndex [24], and then match the companies in WEPP to the companies in MSCI (2018) based on the "bvd_id", which is an identical id for each company and is introduced into WEPP and MSCI (2018) datasets by CorpIndex [24].

The WEPP and MSCI (2018) datasets have 33 281 parent companies and 13 681 companies originally. After matching and enrichment by CorpIndex [24], the WEPP dataset still has 33 281 companies, while the MSCI (2018) dataset has 13 680 companies, losing 1 company without matching. The matching result between WEPP and MSCI (2018) in a sample of 1002 companies, with 857 companies have the revenue data.

4.2.2 WEPP scope 1 carbon intensity estimation

To estimate the scope 1 carbon intensity for the companies in the WEPP dataset, we leverage the estimated scope 1 carbon emissions based on the WEPP dataset and the company revenue data from the MSCI (2018) dataset, given by

$$Scope\ 1\ carbon\ intensity = \frac{Scope\ 1\ carbon\ emissions}{Revenue}. \quad (3)$$

4.2.3 Electric power generation company definition

The WEPP dataset provides the data of electric power generating units and the ownership structure, but not all the companies that own power plants included in the WEPP dataset belong to the energy sector and are electric power generation companies. For example, many paper mills have combined heat and power plants that use purchased natural gas or coal, as well as black liquor produced in their mills, to process heat and generate electricity, and some companies generate electricity through their own on-site solar photovoltaic systems [38]. Hence, defining the electric power generation companies is needed in the WEPP dataset.

For the electric power generation companies, their main business activity is the electric power generation, and almost half the electricity produced in thermal power plants using fossil fuels, with fossil fuels (coal, oil, gas) as the predominant energy source for electricity production [16]. Thus, their carbon intensity is higher than companies belonging to other sectors, which can be used to define the relatively pure electric power generation companies in the WEPP dataset.

Since the coverage of the MSCI (2018) dataset, in terms of company and industry, is way larger than the WEPP dataset, the MSCI (2018) dataset is chosen as the base to decide the benchmark of the relatively pure electric power generation companies in the WEPP dataset. According to the MSCI (2018) scope 1 carbon intensity percentile plots, shown in FIG. 1, the 95^{th} percentile of the MSCI (2018) scope 1 carbon intensity is around 1000 tonsCO_2e/mln USD. Applying the Pareto Principle and leveraging the percentile plots, we set the threshold of scope 1 carbon intensity as 600 tonsCO_2e/mln USD to define the relatively pure electric power generation companies, i.e., for companies in the WEPP dataset with carbon intensity equal to or more than 600 tonsCO_2e/mln USD, they are defined as the electric power generation companies. In the MSCI (2018) dataset, more than 80% of total scope 1 carbon intensity is contributed by around 6.7% companies, with scope 1 carbon intensity greater than or equal to 600 tonsCO_2e/mln USD, shown in FIG. 2. Hence, the companies with scope 1 carbon intensity greater than or equal to 600 tonsCO_2e/mln USD can be defined as relatively pure electric power generation companies. This dataset that only contains the defined electric power generation companies is used as the basis for further analysis.

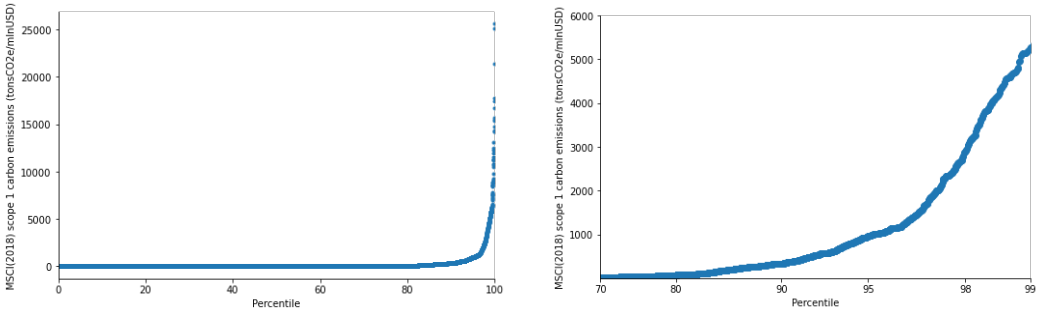

FIG. 1: Percentile plots of the MSCI (2018) scope 1 carbon intensity. The left panel shows the 1st to 99th percentile of the MSCI (2018) scope 1 carbon intensity, indicating that the MSCI (2018) scope 1 carbon intensity increases dramatically from around the 90th percentile. The right panel shows the 70th to 99th percentile of the MSCI (2018) scope 1 carbon intensity, giving a detailed look at the turning point in the percentile plot. The percentile plots of the MSCI (2018) scope 1 carbon intensity could be used to define the original range of the scope 1 carbon intensity threshold for the relatively pure electric power generation companies.

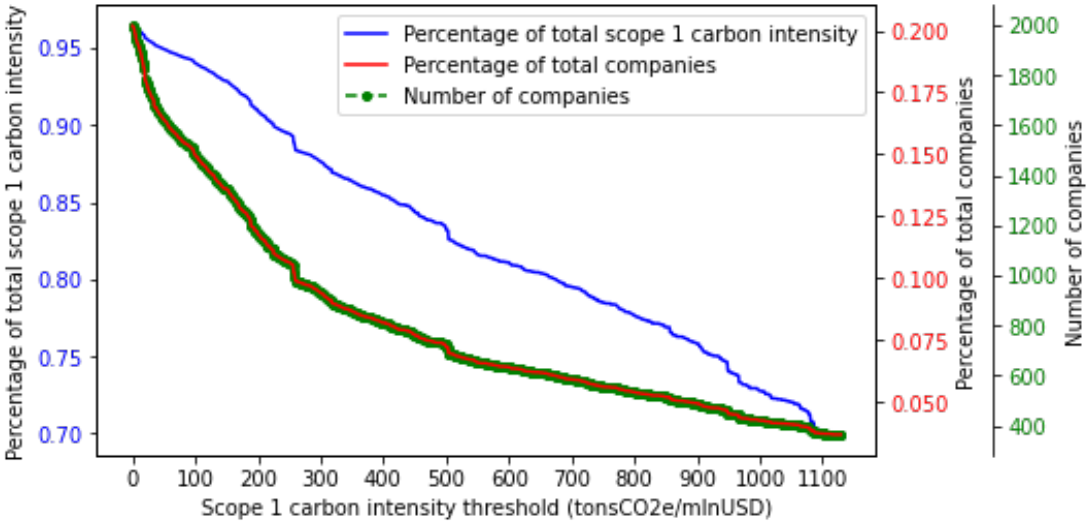

FIG. 2: The scope 1 carbon intensity from 0 to 1100 tonsCO_2e/mln USD to define the threshold for the relatively pure electric power generation companies. The blue line represents the percentage of the total scope 1 carbon intensity contributed by the companies with the scope 1 carbon intensity larger than or equal to the threshold. The red line represents the number of companies, with the scope 1 carbon intensity larger than or equal to the threshold, accounts for the total number of companies in the MSCI (2018) dataset universe. The green dashed-dotted line represents the number of companies with scope 1 carbon intensity larger than or equal to the threshold. For instance, if the scope 1 carbon intensity threshold is set as 600 tonsCO_2e/mln USD, it means that companies with the scope 1 carbon intensity larger than or equal to 600 tonsCO_2e/mln USD are defined as relatively pure electric power generation companies. Their total scope 1 carbon intensity contributes around 83% of the total scope 1 carbon intensity produced by all the companies in the MSCI (2018) dataset universe. In the meanwhile, the number of these companies is around 600, which accounts for around 6.5% of the total number of companies.

4.2.4 Scope 1 carbon emissions difference calculation

For the data quality validation of the MSCI (2018) dataset leveraging the WEPP dataset, we only focus on the scope 1 carbon emissions instead of the scope 1 carbon intensity. Since we use the revenue data from the MSCI (2018) dataset to estimate the scope 1 carbon intensity based on the WEPP dataset, the pattern and thus the analysis of scope 1 carbon emissions are the same as it for the scope 1 carbon intensity, only with a different scale, caused by the revenue.

We calculate the difference between the estimated scope 1 carbon emissions based on the WEPP dataset and the one provided by the MSCI (2018) dataset and use it as the base of further analysis, instead of the absolute difference or relative difference. For absolute difference, the positive and negative differences are ignored so that it's hard to detect if it's overestimation or underestimation and accordingly the environmental impact due to the inaccurate data is overestimated. For relative difference, it may cover up the big difference, i.e., the relative difference can be small, even though the difference between the estimated scope 1 carbon emissions based on the WEPP dataset and the one provided by the MSCI (2018) dataset is large with huge scope 1 carbon emissions provided by MSCI (2018). The scope 1 carbon emissions difference between the one estimated based on the WEPP dataset and the one provided by the MSCI (2018) dataset is calculated based on is given by

$$Scope\ 1\ carbon\ emissions\ difference =$$

$$WEPP\ scope\ 1\ carbon\ emissions - MSCI\ (2018)\ scope\ 1\ carbon\ emissions. \quad (4)$$

Estimated scope 1 carbon emissions based on WEPP is supposed to be greater than or equal to the scope 1 carbon emissions provided by MSCI because the WEPP dataset only provides unit capacity value data and thus the estimated carbon emissions

cover part of the total scope 1 carbon emissions, i.e., the part from the on-campus stationary combustion of fossil fuels.

4.2.5 Statistical tests

To determine whether a relationship is existing between the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018), and if so, how significant or how strong this relationship is between them. We conduct the Spearman's rank correlation test to test if there is an association and the Wilcoxon signed-rank test to test if their distributions are the same or not.

4.2.5.1 Spearman's rank correlation test

Correlation is measured by the correlation coefficient statistically, and there are different types of correlation coefficients. We investigate the data characteristics to decide the most suitable correlation coefficient for our analysis.

The most appropriate coefficient, in this case, is Spearman's rank correlation coefficient due to the following four reasons: (1) outliers, (2) nonnormality of variables, (3) nonlinearity, (4) heteroskedasticity. The scatter plot (FIG. 3) and the distribution plots (FIG. 4 and FIG. 5) of the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018) does not seem compatible with a bivariate normal distribution, and the relationship appears to be monotonic but nonlinear. Besides, there are some relevant outliers. Spearman's rank correlation coefficient, which is robust when outliers are present, can be utilized for the analysis of the monotonic association between such data [39].

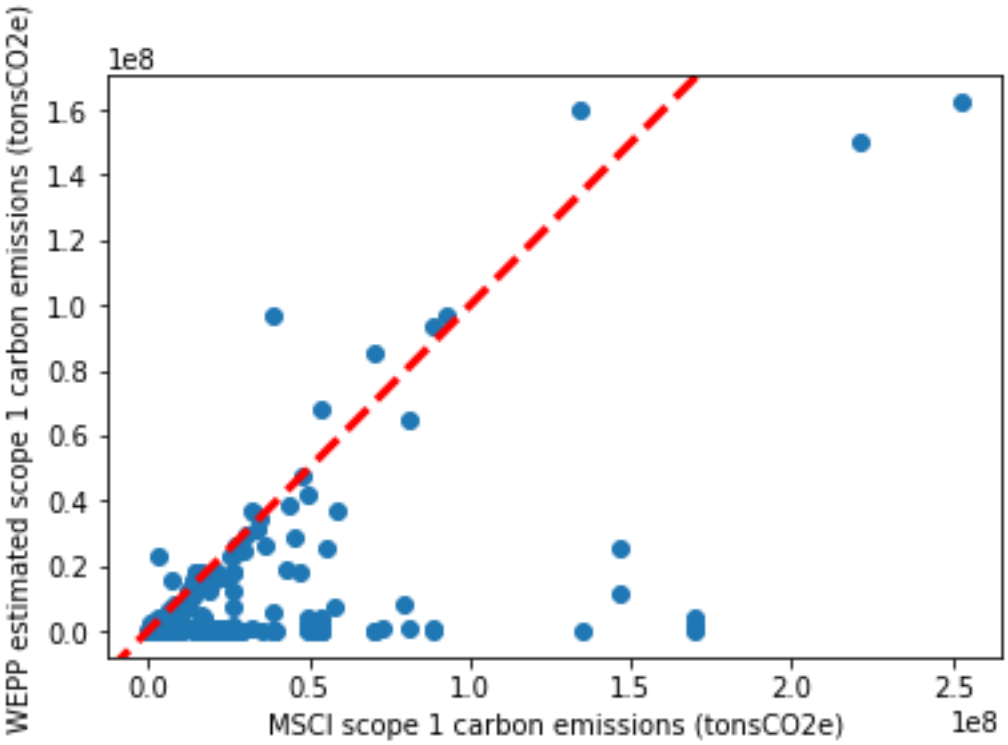

FIG. 3: The scatter plot between the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018). The red line represents the 45-degree diagonal line. Theoretically, all the data points should be in/ under the 45-degree diagonal line as the scope 1 carbon emissions estimated based on the WEPP dataset only cover part of the total scope 1 carbon emissions, i.e., the scope 1 carbon emissions provided by the MSCI (2018) dataset. The scatter plot shows the relationship between the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018) is not linear nor homoscedastic, and there are potential outliers, indicating that the most appropriate coefficient, in this case, is the Spearman's rank correlation coefficient, instead of the Pearson correlation coefficient, which is used frequently for correlation analysis.

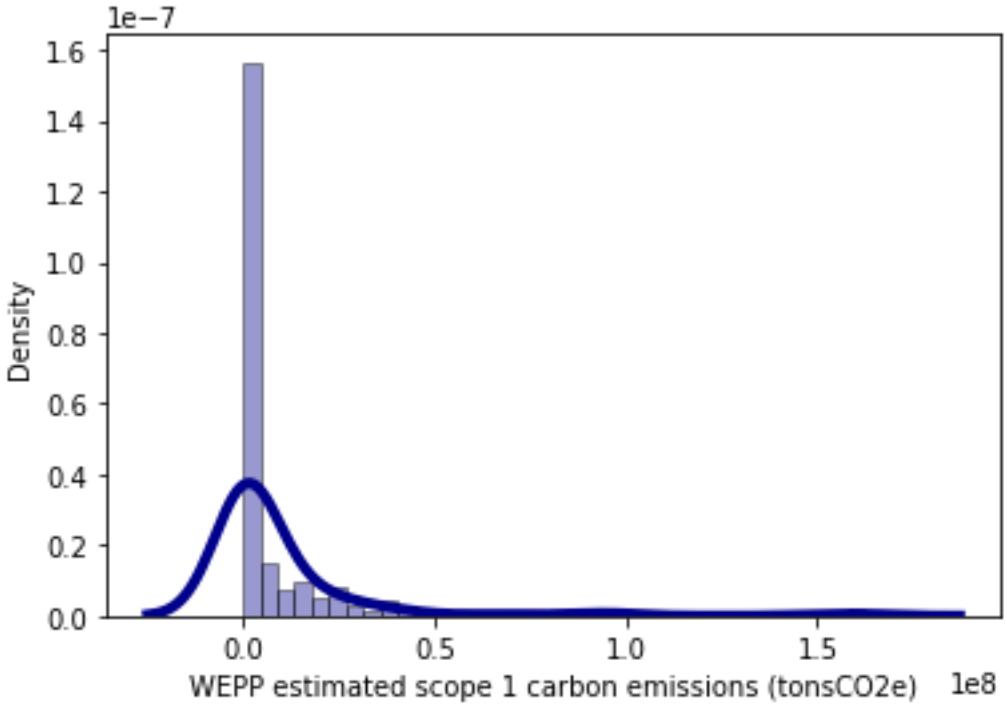

FIG. 4: The distribution plot of the estimated scope 1 carbon emissions based on the WEPP dataset, which is not normally distributed and highly right-skewed or positive-skewed.

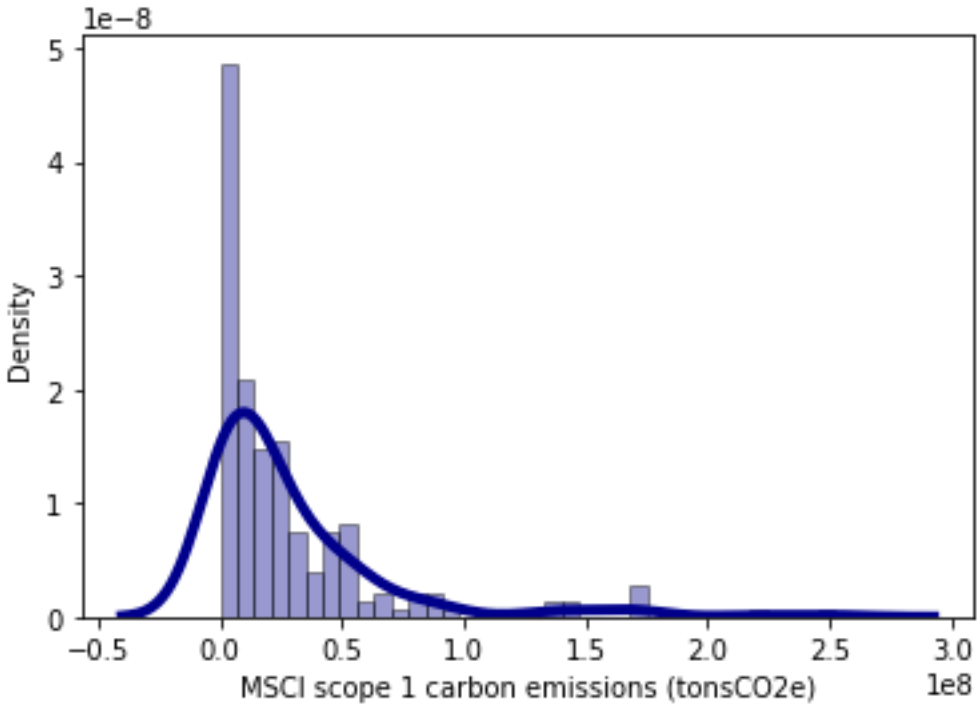

FIG. 5: The distribution plot of the MSCI (2018) scope 1 carbon emissions, which is not normally distributed, which is not normally distributed and highly right-skewed or positive-skewed.

The Spearman's rank correlation coefficient is given by

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}, \tag{5}$$

where ρ is Spearman's rank correlation coefficient, d_i is the difference between the ranks of corresponding variables, and n is the number of observations.

Several labeling systems have been suggested to roughly categorize correlation coefficients (in the absolute values) into different descriptive categories, e.g., "weak," "moderate," or "strong" relationship. However, the cutoff points are arbitrary and inconsistent and should be used judiciously depending on the application's needs. Normally most researchers would probably agree that a coefficient of > 0.90 is a very strong relationship and a coefficient between 0.70 to 0.90 indicates a strong relationship, while the strength for values < 0.70 are disputable [39, 40, 41].

In addition, the coefficient of determination (R^2) can be utilized to better evaluate and interpret the correlation coefficient and assess its practical significance. It's calculated by squaring the correlation coefficient and is defined as the percent of the variation in the dependent variable's values that can be "explained" by changes in the independent variable's value, indicating how much variation in one variable is linked to variation in the other [42].

The Spearman's rank correlation coefficient, ρ, tells us about the strength and direction of the monotonic relationship between the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018). However, the reliability of the monotonic relationship also depends on how many observed data points are in the sample. We need to look at both the value of Spearman's rank correlation coefficient ρ and the number of observations n, together. Thus, we perform a hypothesis test of the "significance of the correlation coefficient" to decide

whether the monotonic relationship in the sample data is strong enough to use the relationship in the population.

Hypothesis tests can be used to determine the statistical significance of the results and to estimate the strength of the relationship in the population from which the data is sampled. The null hypothesis (H0) and the alternative hypothesis (H1) for Spearman's rank correlation test in this case are:

H0: There is no monotonic association between the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018) in the population.

H1: There is a monotonic association between the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018) in the population.

The significance level (denoted as α or alpha) is set as 0.05. A significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference.

4.2.5.2 Wilcoxon signed-rank test

To determine whether the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018) are drawn from the same population distributions, we use the Wilcoxon signed-rank test to conduct this statistical hypothesis test.

The Wilcoxon signed-rank test is chosen due to the following three reasons: (1) the scope 1 carbon emissions data estimated based on WEPP and the one provided by MSCI (2018) is uniquely matched, (2) both do not follow a Gaussian distribution, as it shown in FIG. 4 and FIG. 5, (3) there are some potential outliers. The Wilcoxon signed-rank test is much more robust against outliers and heavy tail distributions.

The Wilcoxon signed-rank test statistic is given by

$$W = \sum_{i=1}^{N}[\mathrm{sgn}(X_i - Y_i) \cdot R_i], \quad (6)$$

where W is the Wilcoxon signed-rank test statistic, X_i, Y_i are the paired data samples from two distributions, N is the sample size excluding pairs where $X_i = Y_i$, sgn denotes the sign function that is $\mathrm{sgn}(x) = 1$ if $x > 0$ and $\mathrm{sgn}(x) = -1$ if $x < 0$, and R_i is the rank of $|X_i - Y_i|$ so that $0 < |X_{R_1} - Y_{R_1}| < |X_{R_2} - Y_{R_2}| < \cdots < |X_{R_n} - Y_{R_n}|$.

The null hypothesis (H0) and the alternative hypothesis (H1) for the Wilcoxon signed-rank test in this case are:

H0: the distributions of the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018) are equal, i.e., both are drawn from a population with the same distribution, and therefore the same population parameters, such as mean or median.

H1: the distributions of the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018) are not equal.

The significance level (denoted as α or alpha) of the Wilcoxon signed-rank test is also set as 0.05.

4.2.6 Analysis for the scope 1 carbon emissions difference

To figure out the potential reasons behind the scope 1 carbon emissions difference between the one estimated based on WEPP and the one provided by MSCI (2018) or if there is any pattern of the difference, we conduct the correlation analysis for the quantitative factors, and the distribution analysis and Kruskal-Wallis H test for each qualitative factor in the dataset.

For the quantitative factors, we first draw the histograms, the scatter plots, and the kernel density estimate (KDE) plots and then calculate Spearman's rank correlation coefficients to find the potential associations. The histograms, the scatter plots, and the KDE plots, can be used to identify trends for follow-up analysis. The histogram and the KDE plots show the distribution of a single variable while the

scatter plots show the relationship (or lack thereof) between two variables. The scatter plots and the Spearman's rank correlation coefficients between the scope 1 carbon emissions difference (tonsCO_2e) and the rest of the quantitative factors, are shown Appendix A. Besides, we calculate Spearman's rank correlation coefficients (ρ), chosen due to the data characteristics shown in the histograms, the scatter plots, and the KDE plots between different quantitative variables (refer to Appendix A).

Leveraging different labeling systems of categorizing correlation coefficients, we define the labeling system in our case, shown in Table 4, used to interpret Spearman's rank correlation coefficient in our analysis.

Table 4: Interpretation of the Spearman's rank correlation coefficient.

The absolute magnitude of the observed correlation coefficient	Interpretation
0.00 - 0.20	Negligible correlation
0.20 - 0.39	Weak correlation
0.40 - 0.69	Moderate correlation
0.70 - 0.89	Strong correlation
0.90 - 1.00	Very strong correlation

For the qualitative factors, we categorize the available qualitative factors in the dataset into four categories: data sources, calculation or estimation methods, industry classifications, and company's locations. For each category, we select the most representative factor for the analysis, i.e., for the industry classifications, we use the Global Industry Classification Standard (GICS) sub-industry level, and for the company's locations, we only focus on the country level rather than the states level, city level, etc.

Since the estimated scope 1 carbon emissions based on WEPP is supposed to be greater than or equal to the scope 1 carbon emissions provided by MSCI, we divide the company distribution analysis into three parts for each qualitative category, one

part is the scope 1 carbon emissions difference larger than 0, another part is the scope 1 carbon emissions difference greater than or equal to 0, and the third part includes all the scope 1 carbon emissions difference.

To test whether samples in each qualitative factor originate from the same distribution, for example, if different data sources lead to different levels of the scope 1 carbon emissions difference, we perform a Kruskal-Wallis test to determine if the median of the scope 1 carbon emissions difference is the same across the different groups of each qualitative factor. The Kruskal-Wallis test is chosen due to the independence of each data sample, the number of observations, and the difference in the data samples size.

The Kruskal-Wallis test statistic is given by

$$H = (N-1)\frac{\sum_{i=1}^{g} n_i(\bar{r}_{i\cdot}-\bar{r})^2}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(r_{ij}-\bar{r})^2}, \quad (7)$$

where H is the Kruskal-Wallis test statistic, N is the total number of observations across all groups, g is the number of groups, n_i is the number of observations in group i, r_{ij} denotes the rank (among all observations) of observation j from group i, $\bar{r}_{i\cdot} = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}$ is the average rank of all observations in group i, and $\bar{r} = \frac{1}{2}(N+1)$ is the average of all the r_{ij}.

The Kruskal-Wallis test uses the following null hypothesis (H0) and alternative hypothesis (H1):

H0: The median is equal across all groups.

H1: The median is not equal across all groups.

The significance level (denoted as α or alpha) for the Kruskal-Wallis test is set as 0.05.

4.3 Consistency between CDP and MSCI (2021)

MSCI ESG Research uses the carbon emissions data reported through CDP when reported data is not available through direct corporate disclosure [4]. Hence, the CDP dataset could also be used to validate the accuracy of the MSCI (2021) dataset and examine the consistency between CDP and MSCI (2021), both for the part that is directly quoted from the CDP dataset and the part that is collected from corporate sources and estimated by MSCI (2021). To make CDP and MSCI (2021) datasets more comparable, the MSCI (2021) dataset is selected for this part of the analysis due to the CDP data availability, only the CDP dataset from 2021 is available.

4.3.1 Data preparation

4.3.1.1 MSCI (2021) data preparation

MSCI provides scope 1 carbon emissions, scope 1 and 2 carbon emissions, and scope 1 and 2 carbon intensity, thus we calculate scope 2 carbon emissions, revenue, scope 1 carbon intensity, and scope 2 carbon intensity, which are given by

$$Scope\ 2\ carbon\ emissions = Scope\ 1\ and\ 2\ carbon\ emissions - Scope\ 1\ carbon\ emissions, \quad (8)$$

$$Revenue = \frac{Scope\ 1\ and\ 2\ carbon\ emissions}{Scope\ 1\ and\ 2\ carbon\ intensity}, \quad (9)$$

$$Scope\ 1\ carbon\ intensity = \frac{Scope\ 1\ carbon\ emissions}{Revenue}, \quad (10)$$

$$Scope\ 2\ carbon\ intensity = \frac{Scope\ 2\ carbon\ emissions}{Revenue}. \quad (11)$$

4.3.1.2 CDP data preparation

CDP provides scope 1 carbon emissions, scope 2 carbon emissions, and scope 1 and 2 carbon intensity, thus we calculate the scope 1 and 2 carbon emissions, revenue, scope 1 carbon intensity, and scope 2 carbon intensity using Eqs. (8) – (11).

For scope 2 carbon emissions, the CDP dataset provides two different data for each company, one is the reported market-based scope 2 carbon emissions and the other one is the location-based scope 2 carbon emissions. Besides, CDP indicates the scope 2 carbon emissions data used for the scope 1 and 2 carbon intensity calculation for each company in its dataset, which is the one that we use for scope 2 carbon intensity calculation and the further analysis related to scope 2 carbon emissions.

4.3.1.3 CDP and MSCI (2021) datasets matching

Since there are no common columns in the original CDP and MSCI (2021) datasets, we link the separate the CDP and MSCI (2021) datasets together for the same company based on the "bvd_id" introduced by Orbis through CorpIndex [24].

The CDP and MSCI (2021) datasets have 7158 companies and 14 742 companies originally. After matching and enrichment by CorpIndex [24], the CDP and MSCI (2021) datasets still hold the same number of companies. The matching result between CDP and MSCI (2021) in a sample of 4170 companies, and there are 4054 companies in the sample after dropping the missing value.

4.3.2 Statistical tests

We investigate the consistency between CDP and MSCI (2021) based on seven different kinds of data in two different coverages of companies. The seven different kinds of data are scope 1 carbon emissions, scope 2 carbon emissions, scope 1 and 2 carbon emissions, scope 1 carbon intensity, scope 2 carbon intensity, scope 1 and 2 carbon intensity, and the revenue data. And the two company coverages are: (1) all the companies that are both in the CDP and MSCI (2021) dataset, (2) the companies that MSCI (2021) directly quotes the relevant data from the CDP dataset or uses the CDP data as the basis for estimation. For the second situation, there are 534 companies in this sample.

4.3.2.1 Spearman's rank correlation test

We use Spearman's rank correlation coefficient to measure the strength of the association and thus the degree of consistency between CDP and MSCI (2021). It's chosen due to the existing outliers, nonnormality of variables, nonlinearity, and heteroskedasticity, which are shown in Appendix B.

4.3.2.2 Wilcoxon signed-rank test

To examine whether the corresponding data from the CDP and the MSCI (2021) datasets are drawn from the same population distributions, we use the Wilcoxon signed-rank test to conduct this statistical hypothesis test. The Wilcoxon signed-rank test is chosen due to three reasons: (1) observations in the CDP and the MSCI (2021) datasets are independent and identically distributed (iid), (2) observations the CDP and the MSCI (2021) datasets can be ranked, (3) observations across the CDP and the MSCI (2021) datasets are paired.

CHAPTER V. RESULTS

5.1 The results of MSCI (2018) dataset validation leveraging WEPP dataset

Since the scope 1 carbon intensity threshold is set as 600 tonsCO_2e/mln USD to define the relatively pure electric power generation companies, there are 211 companies both in the WEPP dataset and the MSCI (2018) dataset with scope 1 carbon intensity equal to or more than 600 tonsCO_2e/mln USD and the descriptive statistics of this dataset is shown in Table 5, which is used as the basis for the analysis.

Table 5: Descriptive statistics of the defined relatively pure electric power generation companies, with scope 1 carbon intensity larger than or equal to 600 tonsCO_2e/mln USD, that are both in the WEPP dataset and the MSCI (2018) dataset. It includes the counts, means, standard deviations (std.), minimum values, 25th percentile, 50th percentile, 75 percentile, and maximum values.

	Estimated electricity generation based on WEPP (MWh)	Scope 1 carbon emissions (tonsCO_2e)		Revenue (mln USD)	Scope 1 carbon intensity (tonsCO_2e/mln USD)	
		Estimated based on WEPP	MSCI (2018)		Estimated based on WEPP	MSCI (2018)
Count	211	211	211	211	211	211
Mean	2.0×10^7	9.8×10^6	2.8×10^7	1.4×10^4	1.4×10^3	3.1×10^3
Std.	4.5×10^7	2.4×10^7	3.9×10^7	2.0×10^4	2.1×10^4	3.0×10^3
Min	26	0	4.8×10^4	20	0	600
25%	2.5×10^5	5.2×10^4	4.1×10^6	1.9×10^3	8.3	990
50%	1.8×10^6	7.2×10^5	1.5×10^7	6.3×10^3	160	2.3×10^3
75%	1.5×10^7	7.3×10^6	3.5×10^7	1.4×10^4	1.5×10^3	4.2×10^3
Max	2.6×10^8	1.6×10^8	2.5×10^8	8.7×10^4	1.9×10^4	1.8×10^4

Besides, the descriptive statistics of the scope 1 carbon emissions difference are shown in Table 6. Estimated scope 1 carbon emissions based on WEPP is supposed to be greater than or equal to the scope 1 carbon emissions provided by MSCI, however, there are 23 companies with the scope 1 carbon emissions difference larger than 0, i.e., the estimation based on the WEPP dataset for around 11% of all the defined relatively pure electric power generation companies is unexpected.

Table 6: Descriptive statistics of the estimated scope 1 carbon emissions based on WEPP (WEPP scope 1 carbon emissions), the scope 1 carbon emissions in MSCI (2018) dataset (MSCI (2018) scope 1 carbon emissions), the carbon emissions difference between the estimated scope 1 carbon emissions based on the WEPP dataset and the one provided by the MSCI (2018) dataset (Scope 1 carbon emissions difference), the carbon emissions difference that is larger than 0 (Scope 1 carbon emissions difference > 0), and the carbon emissions difference that is less than or equal to 0 (Scope 1 carbon emissions difference <= 0). The descriptive statistics contain the counts, means, standard deviations(std.), minimum values, 25th percentile, 50th percentile, 75th percentile, and maximum values.

	WEPP scope 1 carbon emissions ($tonsCO_2e$)	MSCI (2018) scope 1 carbon emissions ($tonsCO_2e$)	Scope 1 carbon emissions difference ($tonsCO_2e$)	Scope 1 carbon emissions difference > 0 ($tonsCO_2e$)	Scope 1 carbon emissions difference <= 0 ($tonsCO_2e$)
Count	211	211	211	23	188
Mean	1.4×10^3	2.8×10^7	-1.8×10^7	$7.\ 7 \times 10^6$	-2.2×10^7
Std.	2.1×10^3	3.9×10^7	3.3×10^7	1.3×10^7	3.3×10^7
Min	0	4.8×10^4	-1.7×10^8	9.2×10^4	-1.7×10^8
25%	8	4.1×10^6	-2.3×10^7	6.0×10^5	-2.6×10^7
50%	160	1.5×10^7	-5.2×10^6	3.1×10^6	-7.9×10^6
75%	1.5×10^3	3.5×10^7	-1.5×10^6	7.2×10^6	-2.3×10^6
Max	1.9×10^4	2.5×10^8	5.8×10^7	5.8×10^7	-7.7×10^4

As for the correlation analysis, a Spearman's rank correlation coefficient of 0.56 is noted between the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018) in the evaluation of these 211 companies in the relatively pure electric power generation sector. The correlation coefficient of 0.56 corresponds to a coefficient of determination (R^2) of 0.32, suggesting that about 32% of the variability of the scope 1 carbon emissions estimated based on WEPP can be "explained" by the relationship with the one provided by MSCI (2018). As more than 68% of the variability is yet unexplained, there must be one or more other relevant factors that are related to the scope 1 carbon emissions provided by MSCI (2018).

According to the results of the Spearman's rank correlation test (p-value = 8.9×10^{-74}), the scope 1 carbon emissions estimated based on WEPP and the one provided

by MSCI (2018) represent a correlation coefficient ($\rho = 0.56$) which is significantly (p-value < α) different from zero. In other words, the relationship existing between the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018) is statistically significant at $\alpha = 0.05$ with 211 samples.

With regard to the analysis of the population distributions, according to the results of the Wilcoxon signed-rank test (statistic = 1.7×10^{-3}, p-value = 7.6×10^{-27}), we do reject H0, i.e., the difference between the distributions is significantly different, because p-value < α (p-value = 7.6×10^{-27}, $\alpha = 0.05$). Therefore, we do have statistically significant evidence at $\alpha = 0.05$, to show that the distributions of the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018) are not equal, i.e., the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018) are not drawn from a population with the same distribution.

According to the statistical tests above, there is a monotonic association existing between the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018) in the population, but there is a significant difference between the distributions of these two samples.

To identify the possible reasons for the scope 1 carbon emissions difference between the one estimated based on WEPP and the one provided by MSCI (2018), we use both quantitative and qualitative elements in the dataset to conduct the correlation analysis, distribution analysis, and the statistical test as well.

For the quantitative factors, the histograms, the scatter plots, and the KDE plots are shown in Appendix A, and the Spearman's rank correlation coefficients between different quantitative variables are shown in FIG. 6.

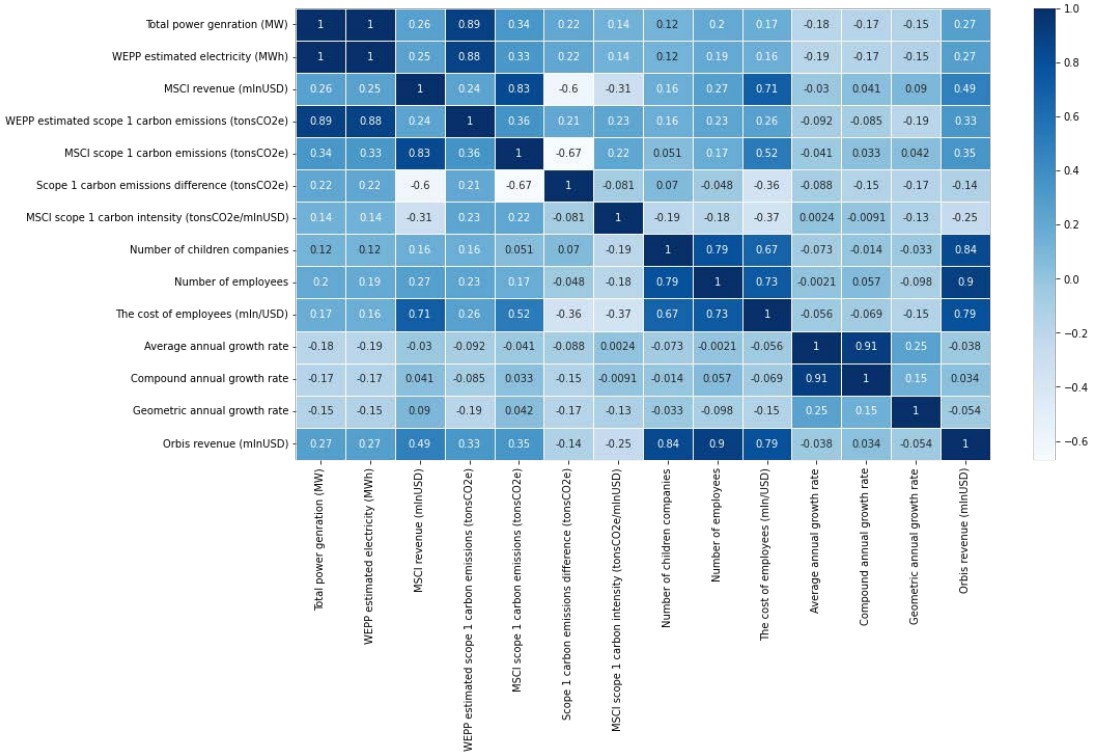

FIG. 6: Spearman's rank correlation coefficients plot of the quantitative factors, i.e., total power generation (MW), WEPP estimated electricity (MWh), MSCI revenue (mlnUSD), WEPP estimated scope 1 carbon emissions (tonsCO_2e), MSCI scope 1 carbon emissions (tonsCO_2e), scope 1 carbon emissions difference (tonsCO_2e), MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), number of children companies, number of employees, the cost of employees (mln/USD), average annual growth rate, compound annual growth rate, geometric annual growth rate, and the Orbis revenue (mln USD).

For the qualitative factors, the company distribution analysis is conducted for four categories (data sources, calculation or estimation methods, industry classifications, and company's locations) under three situations (one part is the scope 1 carbon emissions difference larger than 0, another part is the scope 1 carbon emissions difference greater than or equal to 0, and the third part includes all the scope 1 carbon emissions difference).

The distribution analysis results of data sources and the calculation or estimation methods are shown in Table 7. In total, there are 23 companies (11% of the total companies) with the scope 1 carbon emissions difference larger than 0. There are 8

companies in this part (4% of the total companies, and 35% of this part) source from the Corporate Social Responsibility Reports (SR), and 17 companies' (8% of the total companies, and 74% of this part) data is gathered through reporting. The distribution analysis results of industry classifications, the Global Industry Classification Standard (GICS) sub-industry level, are shown in FIGs. 7-9. The five biggest compositions of the defined relatively pure electric power generation companies are in the electric utilities, multi-utilities, steel, construction materials, and independent power producer & energy traders GICS sub-industry, indicating that our definition for the relatively pure electric power generation companies is accurate. The distribution analysis results of the company's locations, the country level, are shown in FIGs. 10-12.

To investigate if the different groups in each qualitative factor trigger different levels of the scope 1 carbon emissions difference, the Kruskal-Wallis test is conducted, with the results shown in Table 8.

Table 7: The distribution of the MSCI (2018) data sources and the methods used for the defined relatively pure electric power generation companies. The distribution is shown for three parts according to the scope 1 carbon emissions difference between the one estimated based on WEPP and the one provided by MSCI (2018), one part is the scope 1 carbon emissions difference large than 0 (Scope 1 carbon emissions difference > 0), another part is the scope 1 carbon emissions difference no greater than 0 (Scope 1 carbon emissions difference <= 0), and the third part is all the scope 1 carbon emissions difference (Total). There are both the number of companies in each category and the corresponding percentage of each category in the defined relatively pure electric power generation company universe. These statistics are given for: (1) nine data sources, including Annual Reports (AR), Corporate Social Responsibility Reports (SR), Carbon Disclosure Project (CDP), Annual Reports and Corporate Social Responsibility Reports (AR + SR), Annual Reports or Corporate Social Responsibility Reports, and Carbon Disclosure Project (AR / SR + CDP), third party, website, filings, and unknown sources (NaN), (2) eight methods, including quotation of reported data (Reported), estimation using the production model (E. Production Data), estimation using the company-specific intensity model (E.CSI), and estimation using industry segment-specific intensity model with low confidence level (E. Segmt-Low), moderately low confidence level (E.Segmt-Moderate Low), moderate confidence level (E. Segmt-Moderate), moderately high confidence level (E. Segmt-Moderate High), and high confidence level (E. Segmt- High).

Category		Scope 1 carbon emissions difference > 0	Scope 1 carbon emissions difference <= 0	Total
Data sources	AR	3 (1.4%)	35 (16.6%)	38 (18.0%)
	SR	8 (3.8%)	65 (30.8%)	73 (34.6%)
	CDP	3 (1.4%)	37 (17.5%)	40 (19.0%)
	AR + SR	1 (0.5%)	2 (1.0%)	3 (1.4%)
	AR / SR + CDP	0 (0.0%)	1 (0.5%)	1 (0.5%)
	Third party	1 (0.5%)	2 (1.0%)	3 (1.4%)
	Website	1 (0.5%)	0 (0.0%)	1 (0.5%)
	Filings	0 (0.0%)	2 (1.0%)	2 (1.0%)
	NaN	6 (2.8%)	44 (20.9%)	50 (23.7%)
	Total	23 (10.9%)	188 (89.1%)	211
Methods	Reported	17 (8.1%)	139 (65.9%)	156 (73.9%)
	E. Production Data	1 (0.5%)	13 (6.2%)	14 (6.6%)
	E. CSI	0 (0.0%)	5 (2.4%)	5 (2.4%)
	E. Segmt-Low	0 (0.0%)	3 (1.4%)	3 (1.4%)
	E. Segmt-Moderate Low	0 (0.0%)	3 (1.4%)	3 (1.4%)
	E. Segmt-Moderate	1 (0.5%)	4 (1.9%)	5 (2.4%)
	E. Segmt-Moderate High	3 (1.4%)	14 (6.6%)	17 (8.1%)
	E. Segmt- High	1 (0.5%)	7 (3.3%)	8 (3.8%)
	Total	23 (10.9%)	188 (89.1%)	211

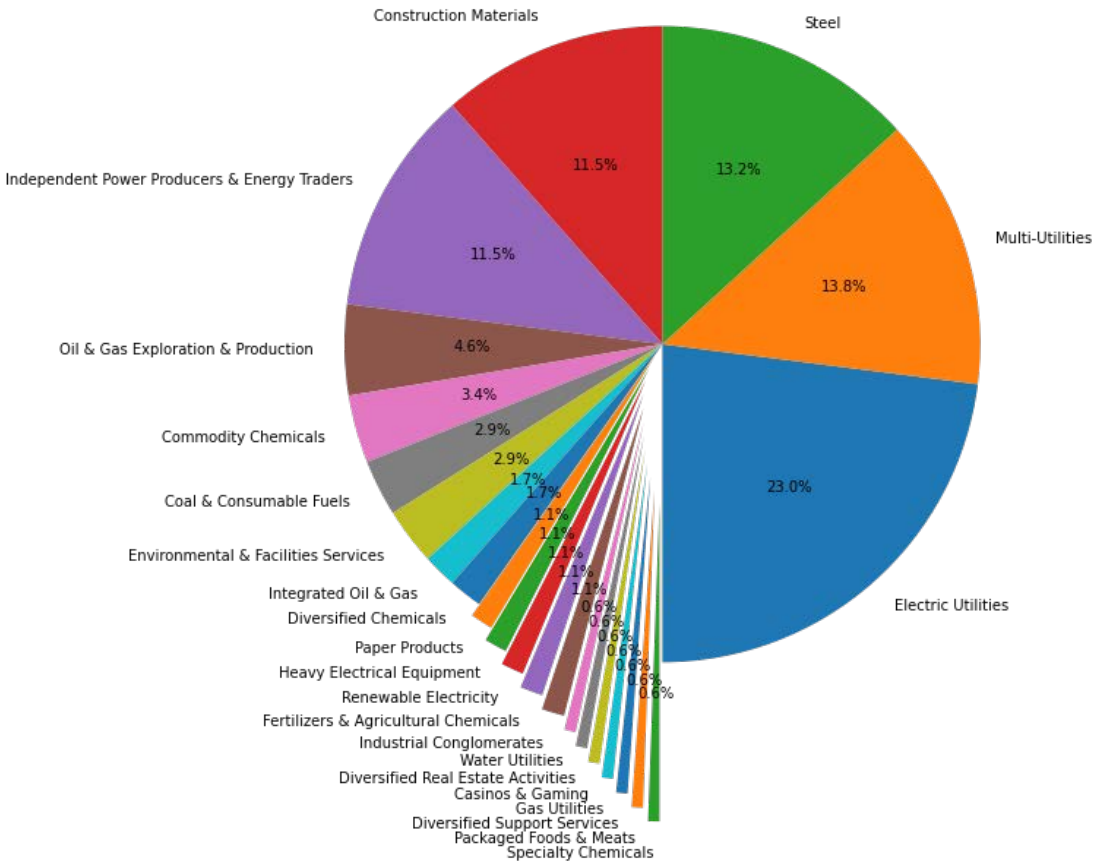

FIG. 7: The distribution of industry classifications, the Global Industry Classification Standard (GICS) sub-industry level, for all the defined relatively pure electric power generation companies, that are 211 companies in total.

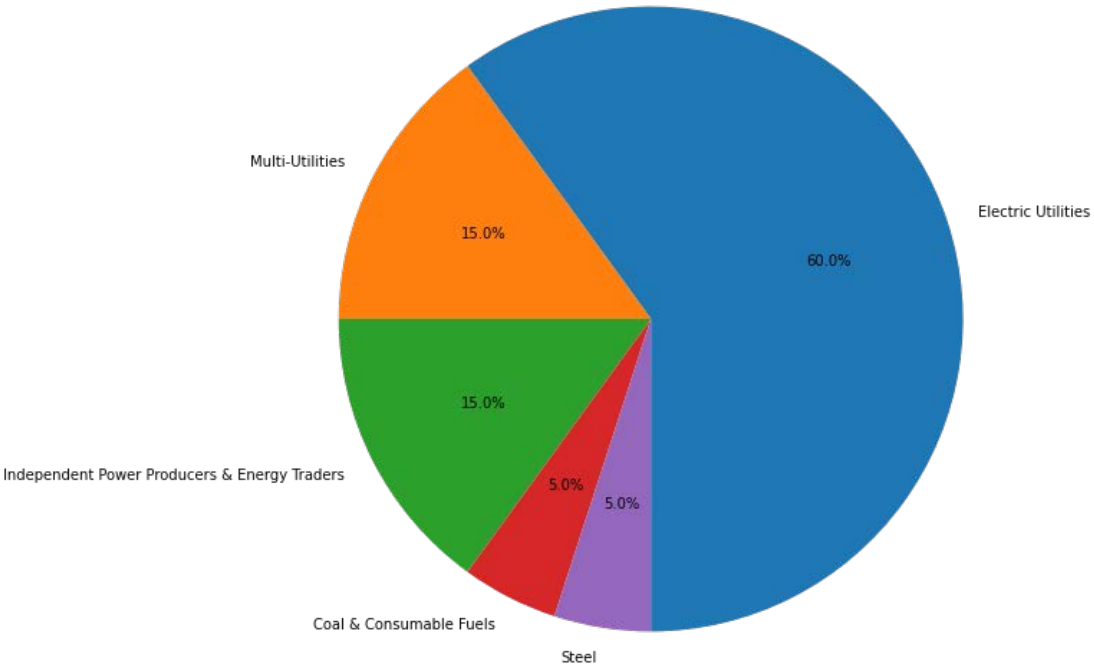

FIG. 8: The distribution of industry classifications, the Global Industry Classification Standard (GICS) sub-industry level, for the defined relatively pure electric power generation companies with the scope 1 carbon emissions difference larger than 0. There are 23 defined relatively pure electric power generation companies with the scope 1 carbon emissions difference larger than 0 in total.

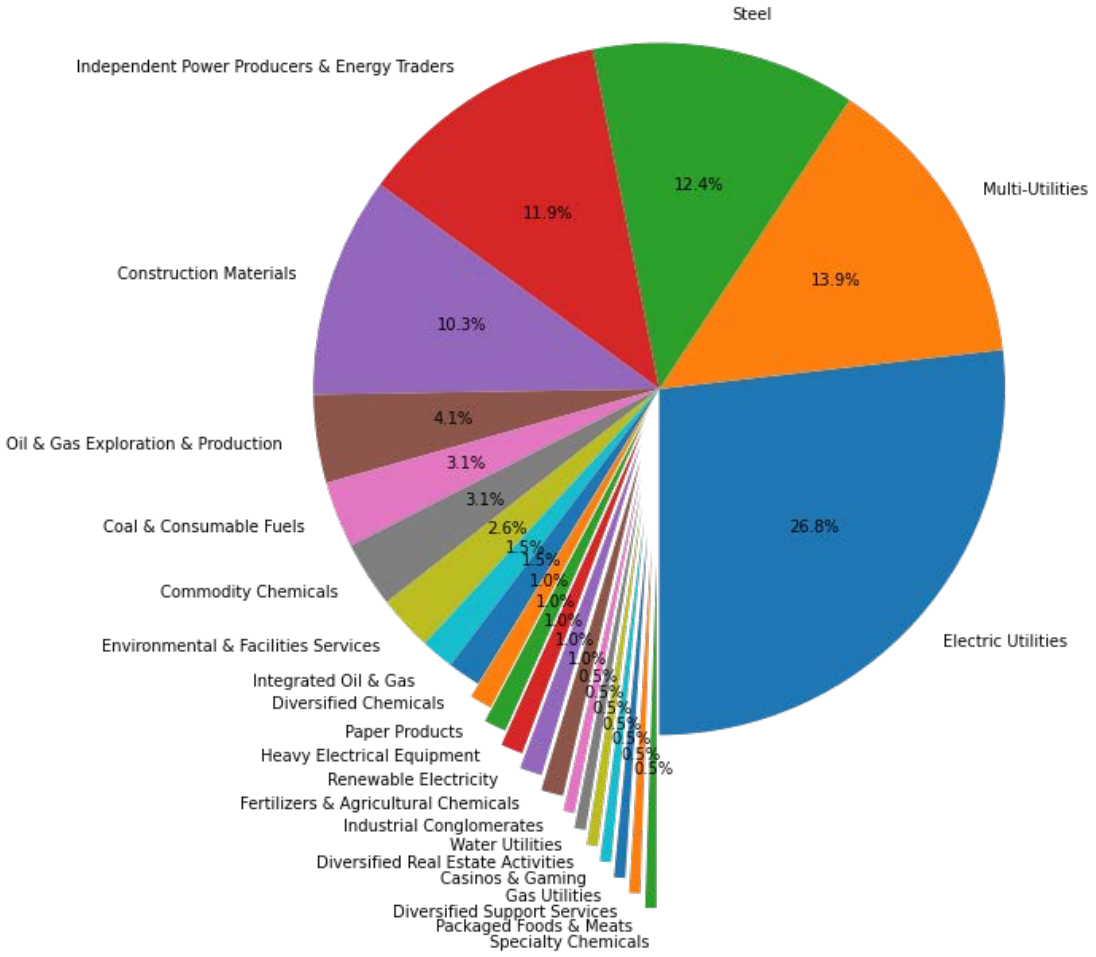

FIG. 9: The distribution of industry classifications, the Global Industry Classification Standard (GICS) sub-industry level, for the defined relatively pure electric power generation companies with the scope 1 carbon emissions difference greater than or equal to 0. There are 188 defined relatively pure electric power generation companies with the scope 1 carbon emissions difference greater than or equal to 0 in total.

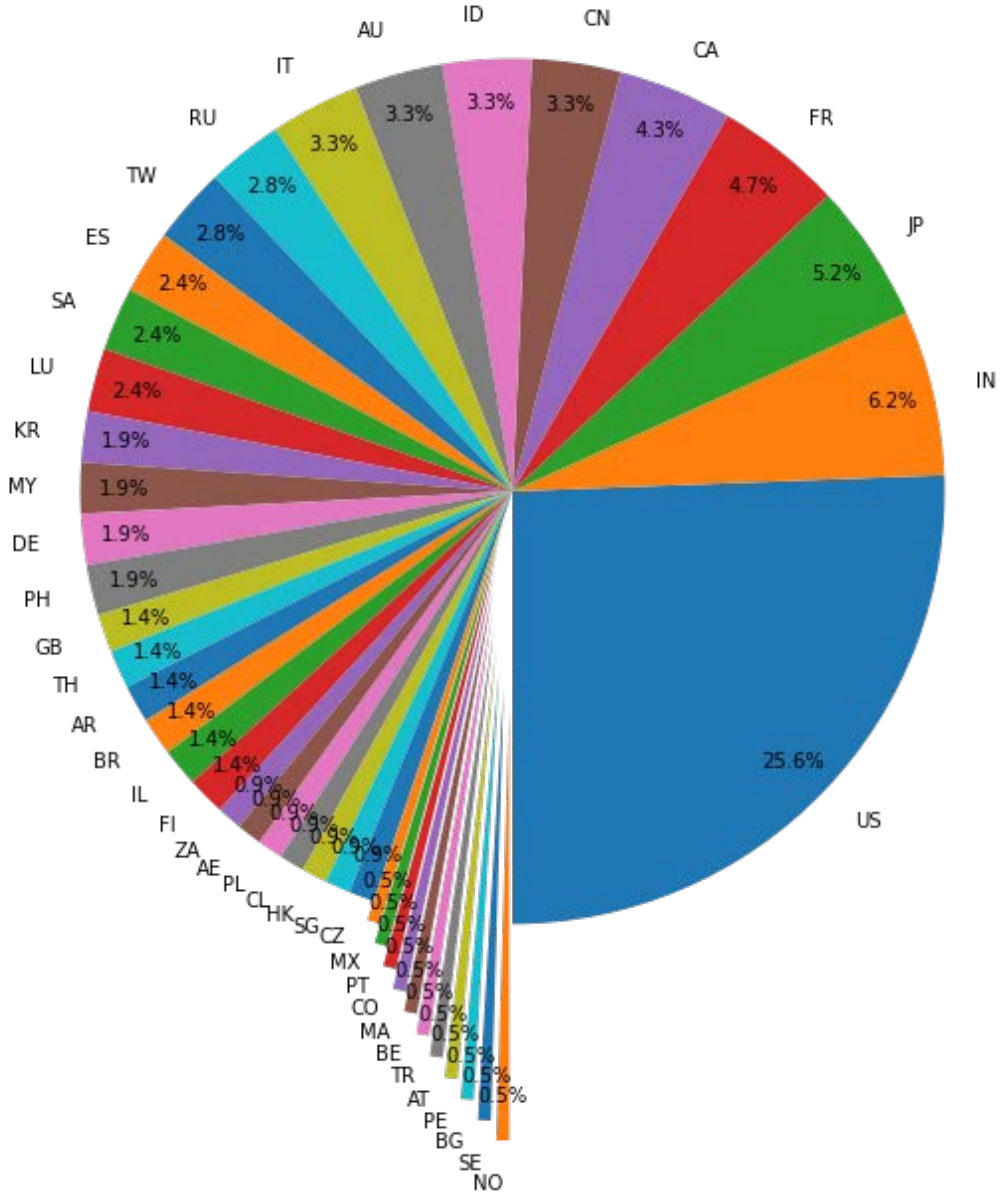

FIG. 10: The distribution of the country, using country abbreviation ISO-3166-1 alpha-2 country code standard [43], that the defined relatively pure electric power generation companies are located. There are 211 defined relatively pure electric power generation companies.

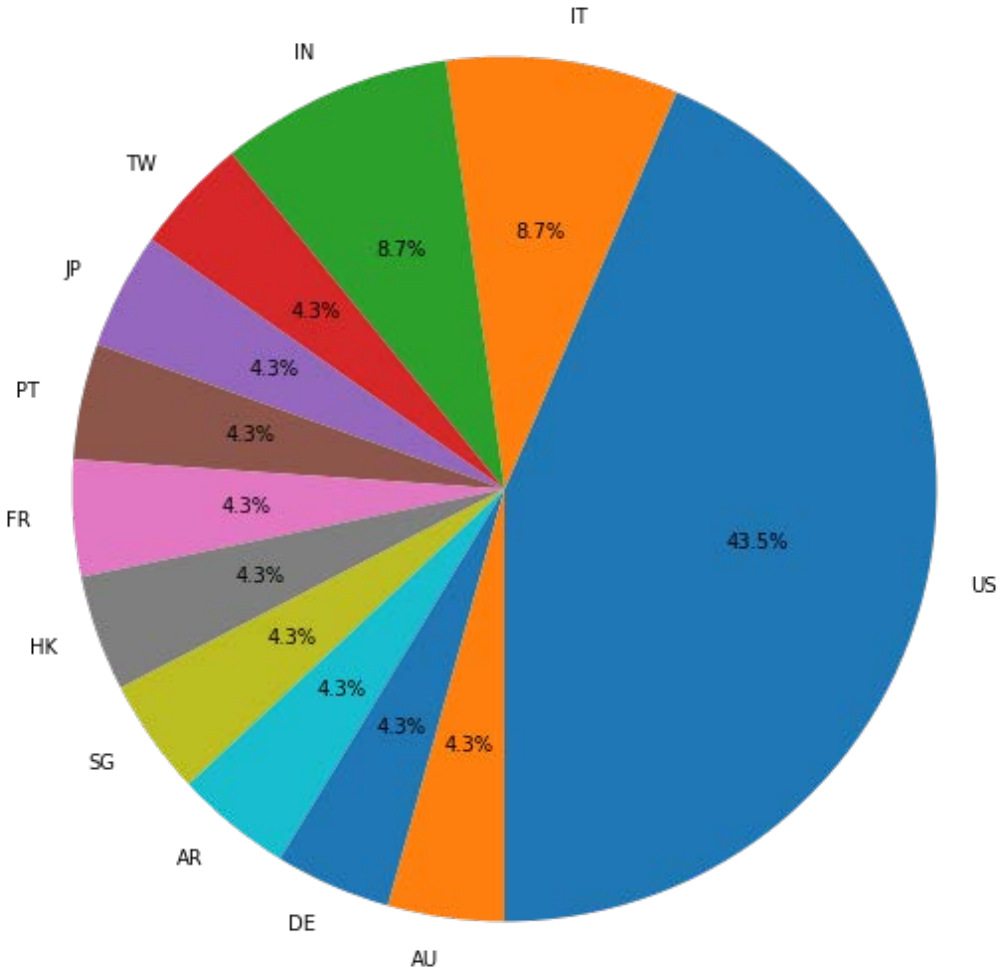

FIG. 11: The distribution of the country, using country abbreviation ISO-3166-1 alpha-2 country code standard [43], that the defined relatively pure electric power generation companies with the scope 1 carbon emissions difference larger than 0 are located. There are 23 defined relatively pure electric power generation companies with the scope 1 carbon emissions difference larger than 0 in total.

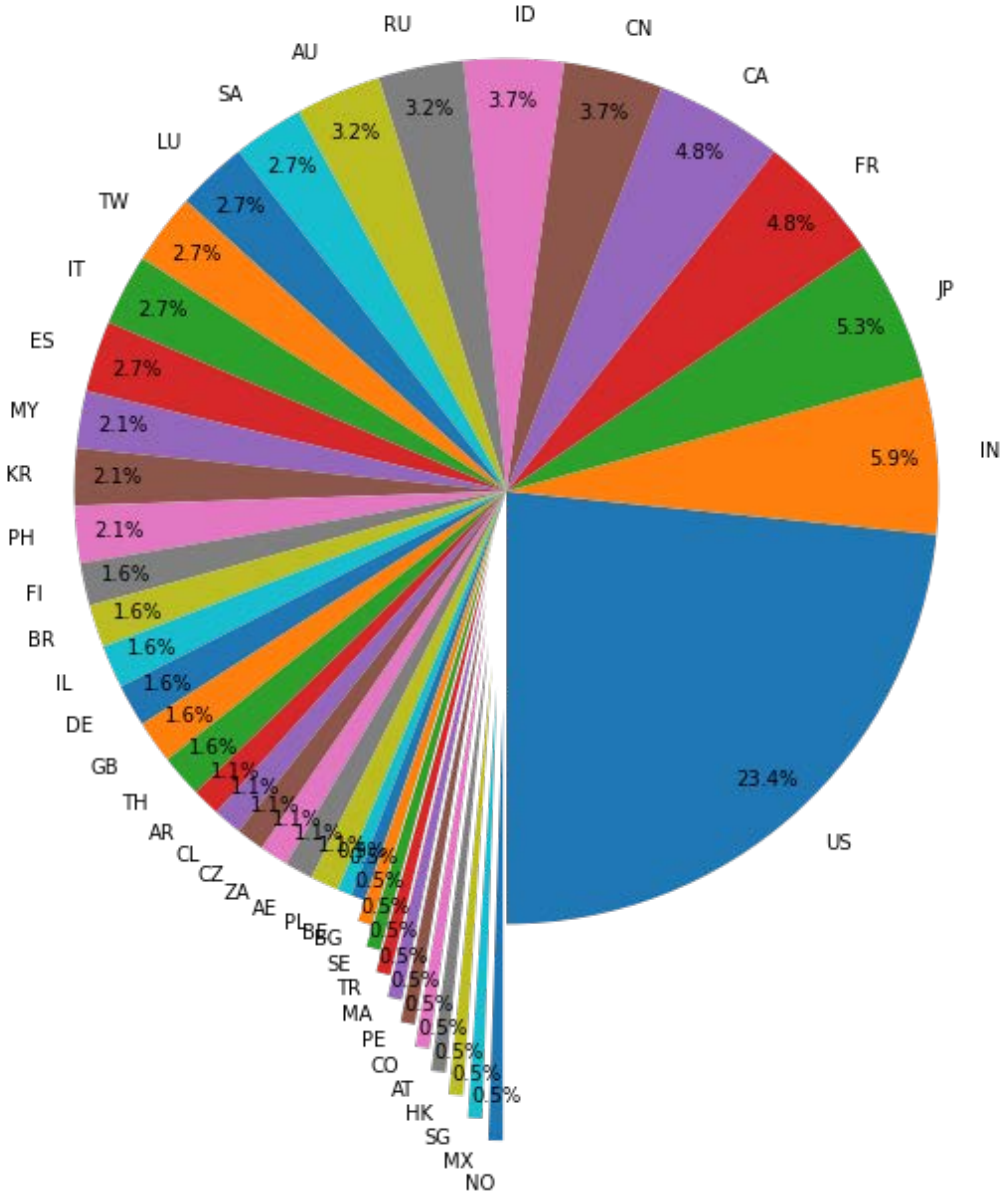

FIG. 12: The distribution of the country, using country abbreviation ISO-3166-1 alpha-2 country code standard [43], that the defined relatively pure electric power generation companies with the scope 1 carbon emissions difference greater than or equal to 0 are located. There are 188 defined relatively pure electric power generation companies with the scope 1 carbon emissions difference larger than 0 in total.

Table 8: Kruskal-Wallis test for the different groups in four qualitative factors: (1) data sources, (2) methods used in calculating and estimating the scope 1 carbon emissions in MSCI (2018), (3) industry classifications, the Global Industry Classification Standard (GICS) sub-industry level, and (4) country, using country abbreviation ISO-3166-1 alpha-2 country code standard [43]. This test aims to figure out whether the different groups of data sources and methods originate from the same distribution, i.e., if these different groups have the same impact on the scope 1 carbon emissions between the one estimated based on WEPP and the one provided by MSCI (2018).

Qualitative factors	Kruskal-Wallis test	
	Statistic	P-value
Data sources	18	0.01
Methods	22	0.00
Industry	20	0.61
Country	49	0.19

5.2 The results of the consistency between CDP and MSCI (2021)

After matching the CDP and MSCI (2021) datasets, there are 4170 companies in this sample, with 4054 companies remain in the sample after dropping the missing value. A sample of 4054 companies with valid data points is used as the basis for the analysis.

To examine the consistency between CDP and MSCI (2021), we conduct both the Spearman's rank correlation test and the Wilcoxon signed-rank test. Two situations are considered for the tests: (1) all the companies that are both in the CDP and MSCI (2021) dataset, (2) the companies that MSCI (2021) directly quotes the relevant data from the CDP dataset or uses the CDP data as the basis for estimation. For the first situation, there are 4054 companies in the sample, while there are 534 companies in the second situation.

The results of Spearman's rank correlation test and the Wilcoxon signed-rank test for the two situations are shown in Table 9.

Table 9: Statistical Tests, including the Spearman's rank correlation test and the Wilcoxon signed-rank test, between the different types of carbon emissions and carbon intensity, and the revenue data in the CDP and MSCI (2021) dataset. The statistical tests are conducted under two situations: (1) all the companies that are both in the CDP and MSCI (2021) dataset, and the results are shown in the upper side, (2) the companies that MSCI (2021) directly quotes the relevant data from the CDP dataset or uses the CDP data as the basis for estimation, and the results are shown in the lower side, i.e., in the parentheses.

Data type		Spearman's rank correlation test		Wilcoxon signed-rank test	
		Statistic	P-value	Statistic	P-value
Carbon emissions	Scope 1	0.88 (0.99)	0.00 (0.00)	1.4×10^{6} (3.6×10^{3})	3.1×10^{-95} (1.9×10^{-10})
	Scope 2	0.79 (0.91)	0.00 (3.1×10^{-200})	2.7×10^{6} (2.3×10^{4})	1.7×10^{-26} (5.7×10^{-16})
	Scope 1 and 2	0.87 (0.97)	0.00 (0.00)	2.4×10^{6} (2.6×10^{4})	5.0×10^{-59} (3.8×10^{-14})
Carbon intensity	Scope 1	0.82 (0.97)	0.00 (0.00)	2.3×10^{6} (4.8×10^{4})	6.7×10^{-134} (5.3×10^{-10})
	Scope 2	0.66 (0.85)	0.00 (1.4×10^{-146})	3.0×10^{6} (4.8×10^{4})	3.2×10^{-45} (4.1×10^{-10})
	Scope 1 and 2	0.79 (0.94)	0.00 (3.0×10^{-242})	2.5×10^{6} (5.4×10^{4})	2.3×10^{-98} (3.0×10^{-6})
Revenue		0.99 (0.99)	0.00 (0.00)	3.3×10^{6} (4.7×10^{4})	2.8×10^{-29} (3.5×10^{-11})

CHAPTER VI. DISCUSSION

6.1 MSCI data quality

According to the results of the Spearman's rank correlation test (statistic = 0.56, p-value = 8.9×10^{-74}) and the Wilcoxon signed-rank test (statistic = 1.7×10^{3}, p-value = 7.6×10^{-27}), there is a monotonic moderately correlation existing between the WEPP estimated scope 1 carbon emissions and the MSCI (2018) scope 1 carbon emissions, and the correlation is statistically significant at $\alpha = 0.05$ with 211 samples. But the distributions of these two samples are significantly different. Based on the calculation results of the Spearman's rank correlation coefficients between different quantitative variables in FIG. 6 and the interpretation of the Spearman's rank correlation coefficient in Table 4, we can know that the scope 1 carbon emissions

difference between the figure estimated based on WEPP and the one provided by MSCI (2018) is moderately correlated with the MSCI (2018) revenue and the scope 1 carbon emissions provided by MSCI (2018). The negative correlation between the scope 1 carbon emissions difference and the MSCI (2018) revenue and the scope 1 carbon emissions provided by MSCI (2018) showing that the defined relatively pure electric power generation companies with higher revenue in the MSCI (2018) dataset and higher scope 1 carbon emissions in the MSCI (2018) dataset tend to be more accurately estimated through the WEPP dataset.

Besides, for most of the defined relatively pure electric power generation companies (188 companies, i.e., 89% of total companies), their scope 1 carbon emissions difference is greater than or equal to 0 as assumed, indicating that most of the scope 1 carbon emissions data in MSCI (2018) for electric power generation companies is in the right direction.

Furthermore, the Kruskal-Wallis test is used to investigate if the distinct groups in each qualitative factor cause various degrees of the scope 1 carbon emissions difference, with the results shown in Table 8. For data sources and methods used for the calculation or estimation of scope 1 carbon emissions, the decision to reject the null hypothesis is made (p-value $< \alpha = 0.05$), indicating that different data sources and methods lead to different calculation and estimation results. While for industry and country, we fail to reject the null hypothesis (p-value $> \alpha = 0.05$) that the median of scope 1 carbon emissions is the same for all the industries and all the countries. Thus, we have sufficient evidence to conclude that the data sources and methods used for calculation and estimation lead to statistically significant differences in the scope 1 carbon emissions difference, nevertheless, the various industry and countries make no difference to scope 1 carbon emissions difference.

Concerning the consistency between MSCI (2021) dataset and the CDP dataset, the results of Spearman's rank correlation test and the Wilcoxon signed-rank test show that the carbon emissions and carbon intensity for scope 1, scope 2, and scope 1 and 2 are highly consistent in two different company coverages, i.e., all the companies that are both in the CDP and MSCI (2021) dataset, and the companies that MSCI (2021) directly cites from the CDP dataset or uses the CDP data as the basis for estimation, indicating that the carbon-related information disclosed by companies publicly is consistent with the one they disclose to CDP. As shown in Table 9, scope 1 and scope 1 and 2 tend to have a higher consistency than scope 2 both for carbon emissions and carbon intensity, no matter considering all the companies nor only considering the companies that MSCI directly cites from CDP or uses CDP for estimation. Comparing the results of these two different company coverages, we could find that the consistency is higher for the part that MSCI (2021) directly cites from CDP or estimation based on CDP than the companies both in MSCI (2021) and CDP dataset. However, the distributions of all the carbon emissions and carbon intensity of MSCI (2021) and CDP are different in both company coverages.

In contrast to the consistency of scope 1 carbon emissions between MSCI (2021) and CDP (ρ = 0.88 or ρ = 0.99), the consistency of between the scope 1 carbon emissions estimated based on WEPP and the one provided by MSCI (2018) is much lower (ρ = 0.56). But the distribution of the MSCI dataset is different from the distribution of both WEPP and CDP datasets.

6.2 Potential reasons for the data discrepancy

Sources of discrepancies for scope 1 carbon emissions between WEPP estimation and the MSCI reporting result from differences in the following three aspects: (1) raw data sources, (2) methodology, including the capacity factors, the

energy sources classification, and the ownership structure, (3) matching quality between different datasets.

The raw data used to compile carbon emissions data may often be different among different data providers. MSCI sends annual surveys to its member companies as the primary method of collecting data. By contrast, WEPP relies primarily on national reports and information from regional agencies. Differences in surveys and collection sources can lead to disparities in the values of reported physical quantities of fuels (such as tonnes of coal or m^3 of natural gas).

The potential discrepancy sources of the methodology are capacity factors, the energy source classification, and ownership structure.

When estimating the carbon emissions using the production model for the electric power generation companies, the capacity factors used in the model vary substantially depending on the plant and fuel source [26]. To convert estimates of physical quantities of electric power generated into carbon emissions values, data providers utilize a conversion factor termed capacity factor. The capacity factor of an energy source is the ratio of a given period's actual electricity generated over a given period to the maximum potential electrical energy output during the same period [44]. Data providers estimate the carbon emissions of each plant or each company based on the capacity factors of energy sources. Capacity factors utilized by data providers are country- and region-specific [44], because the technical constraints, such as availability of the plant, economic reasons, and availability of the energy resource of different plants are often not uniform within any particular country. Country-specific capacity factors utilized by data providers are often different, which has the effect of creating apparent differences in carbon emissions of the same company with plants in various countries where the reported value of physical quantities of energy sources

consumed are identical. For our estimation, we use the average capacity factors for each energy source, while MSCI doesn't disclose the capacity factors used in their calculation. The potential difference of the capacity factors used in the calculation may bring the discrepancy of the carbon emissions estimation.

Aside from capacity factors differences, carbon emissions may differ owing to the energy source classification, i.e., the boundary criteria for which energy sources are included. The inclusion or exclusion of international bunker fuels, modern renewable energy sources, and energy from biomass and wastes are the most evident differences in system boundaries [12].

Further differences in carbon emissions between the one estimated based on the WEPP dataset and the one provided by MSCI may result from the ownership structure of companies utilized for the estimation, especially when it comes to the joint venture. A substantial number of power facilities, notably large nuclear, coal, and hydroelectric plants, are jointly owned. Power companies, as well as other parties such as fuel or manufacturing corporate entities, investment funds and financial institutions, and different national or local government agencies, could be among the owners. Since the WEPP database does not track joint ownership shares and the parent company in the WEPP dataset may also contain listings for two or more companies in the joint venture or other arrangements, we aggregate plant-level and unit-level data to a higher "institutional" level only based on the available information in the WEPP dataset. However, the ownership structure of the corresponding companies in the MSCI dataset is unclear, which may cause the estimation difference to a certain degree.

Beyond all the possible sources of the difference, the company matching quality between WEPP and MSCI datasets may also bring additional disturbance to the whole

analysis. To make parameters of the same company across different datasets comparable, it's necessary to match these datasets sophistically, especially when there is no common aspect to indicate the identity of the companies. In our analysis, the CorpIndex [24], mainly using fuzzy name matching algorithm implemented by Swiss Re, is the tool we leverage for the company matching among different datasets, however, it may result in mismatching when it comes to the joint venture.

CHAPTER VII. CONCLUSIONS AND FUTURE WORK

Our results of the MSCI carbon emissions data validation provide detailed insights on the data quality issues. There is considerable uncertainty inhabited in the MSCI carbon emissions data at least for the relatively pure power generation companies. Given the potential severe climatic consequences and enormous economic and political implications of efforts to reduce carbon emissions, accurate, consistent, and comparable datasets and their associated uncertainties are needed for transparent reference and progress monitoring [9]. Additional attention to the data quality issues in this area from all the stakeholders is needed and special efforts should be taken to improve the data quality.

Different primary data sources and methodology (including capacity factors, the energy sources classification, and the ownership structure) can lead to significantly different results of carbon emissions by the choice of one dataset over another. Thus, stakeholders are suggested to conduct data quality validation and assessment using alternative independent data sources, which can provide a more comprehensive glimpse into what actual carbon emissions might be. For effective independent validation and verification, data collection, and reporting, and the overall transparency must improve among different data providers, providing directly comparable values.

Besides, it's also important to improve the company matching quality across different datasets to ensure an accurate comparison.

Improving the quality and consistency of data in carbon emissions data providers could facilitate the development of more robust mitigation measures to reduce carbon emissions.

REFERENCES

[1] IPCC, "Climate Change 2021: The Physical Science Basis," IPCC, 2021.

[2] "Greenhouse gas emissions," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Greenhouse_gas_emissions.

[3] "ghgprotocol," ghgprotocol, [Online]. Available: https://ghgprotocol.org.

[4] MSCI, *MSCI Carbon Emissions Estimation Methodology.*

[5] United Nations, *The Paris Agreement,* Paris, 2015.

[6] D. Pandey, M. Agrawal and J. S. Pand, "Carbon Footprint: Current Methods of Estimation," *Environmental Monitoring and Assessment,* Vols. 178(1-4), pp. 135-160, July 2011.

[7] A. Grübler, "Trends in Global Emissions: Carbon, Sulfur, and Nitrogen," *Encyclopedia of Global Environmental Change,* vol. 3, pp. 35-53, May 2002.

[8] P. Faria, "Uncertainty and variability in corporate GHG inventories and reporting," in *3rd International Workshop on Uncertainty in Greenhouse Gas Inventories*, Lviv, Ukraine, September 22-24, 2010.

[9] E. Romijn, V. D. Sy, M. Herold, H. Böttcher, R. M. Roman-Cuesta, S. Fritz, D. Schepaschenko, V. Avitabile, D. Gaveau, L. Verchot and C. Martius, "Independent data for transparent monitoring of greenhouse gas emissions from the land use sector – What do stakeholders think and need?," *Environmental Science & Policy,* vol. 85, pp. 101-112, 2018.

[10] V. Kalesnik, M. Wilkens and J. Zink, "Green Data or Greenwashing? Do Corporate Carbon Emissions Data Enable Investors to Mitigate Climate

Change?" 7 January 2021. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3722973.

[11] T. Busch, M. Johnson and T. Pioch, "Corporate carbon performance data: Quo vadis?," *Journal of Industrial Ecology,* 24 April 2020.

[12] J. Macknick, "Energy and CO2 emission data uncertainties," *Carbon Management,* vol. 2, p. 189–205, 2011.

[13] G. Marland, "The U.S. NRC report on monitoring and verification of national greenhouse gas emissions inventories," in *3rd International Workshop on Uncertainty in Greenhouse Gas Inventories*, Lviv, Ukraine, September 22-24, 2010.

[14] M. Ge, J. Friedrich and L. Vigna, "World Resources Institute," 6 February 2020. [Online]. Available: https://www.wri.org/insights/4-charts-explain-greenhouse-gas-emissions-countries-and-sectors.

[15] Agency, United States Environmental Protection, "Global Greenhouse Gas Emissions Data," United States Environmental Protection Agency, [Online]. Available: https://www.epa.gov/ghgemissions/global-greenhouse-gas-emissions-data.

[16] European Environment Agency, "Environmental signals 2000 - Environmental assessment report No 6," European Environment Agency, 2000.

[17] F. Gotzens, H. Heinrichs, J. Hörsch and F. Hofmann, "Performing energy modelling exercises in a transparent way - The issue of data quality," *Energy Strategy Reviews,* vol. 23, pp. 1-12, January 2019.

[18] S&P Global Market Intelligence, "DATA BASE DESCRIPTION AND RESEARCH METHODOLOGY: WORLD ELECTRIC POWER PLANTS

DATA BASE," S&P Global Market Intelligence, Washington, DC 20005 USA, 2017.

[19] "CDP," [Online]. Available: https://www.cdp.net/en.

[20] "Carbon Disclosure Project," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Carbon_Disclosure_Project.

[21] Rolls Royce, April 2019. [Online]. Available: https://www.rolls-royce.com/~/media/Files/R/Rolls-Royce/documents/sustainability/Supplier-docs/CDP.pdf.

[22] E. Stanny, "Reliability and Comparability of GHG Disclosures to the CDP by US Electric Utilities," *Social and Environmental Accountability,* vol. 38:2, pp. 111-130, 2018.

[23] "Bureau van Dijk," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Bureau_van_Dijk.

[24] Swiss Re, *CIX,* Zurich: Swiss Re, 2021.

[25] World Resources Institute, "GREENHOUSE GAS Protocol," [Online]. Available: https://ghgprotocol.org/corporate-standard.

[26] Duke Energy Nuclear Education, "Capacity Factor – A Measure of Reliability," Duke Energy Nuclear Information Center (NIC), 18 February 2015. [Online]. Available: https://nuclear.duke-energy.com/2015/02/18/capacity-factor-a-measure-of-reliability.

[27] "Capacity factor," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Capacity_factor.

[28] C. Shearer, L. Myllyvirta, A. Yu, G. Aitken, N. Mathew-Shah, G. Dallos and T. Nace, "Boom and Bust 2020: TRACKING THE GLOBAL COAL PLANT

PIPELINE," Global Energy Monitor, Greenpeace International, CREA, and Sierra Club, 2020.

[29] A. Kwon, "U.S. Energy Information Administration," U.S. Energy Information Administration, 8 September 2015. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=22832.

[30] M. Mueller, "Office of Nuclear Energy," U.S. Department of Energy, 1 May 2020. [Online]. Available: https://www.energy.gov/ne/articles/what-generation-capacity#:~:text=The%20Capacity%20Factor&text=Capacity%20factors%20allow%20energy%20buffs,power%20all%20of%20the%20time.

[31] IEA-ETSAP and IRENA, "Hydropower: Technology Brief," International Renewable Energy Agency (IRENA), 2015.

[32] IRENA, "Future of Solar Photovoltaic: Deployment, investment, technology, grid integration and socio-economic aspects (A Global Energy Transformation: paper)," International Renewable Energy Agency, Abu Dhabi, 2019.

[33] IRENA, "Future of wind: Deployment, investment, technology, grid integration and socio-economic aspects (A Global Energy Transformation paper)," International Renewable Energy Agency, Abu Dhabi, 2019.

[34] World Nuclear Association, "World Nuclear Performance Report 2018," World Nuclear Association, 2018.

[35] IRENA, "Geothermal Power: Technology Brief," International Renewable Energy Agency, Abu Dhabi, 2017.

[36] Y. Shibata, "Economic Analysis of Hydrogen Production from Variable Renewables," *IEEJ Energy Journal,* vol. 10, no. 2, pp. 26-46, 2015.

[37] B. Alves, "Load factor of electricity from waste energy in the UK 2010-2019 Published by Bruna Alves, Jul 5, 2021, The load factor for electricity generation from waste energy in the United Kingdom has fluctuated since 2010. In 2019, the load factor of energy from," Statista, 5 July 2021. [Online]. Available: https://www.statista.com/statistics/555725/energy-from-waste-electricity-load-factor-uk/.

[38] EIA, "Use of energy explained: Energy use in industry," U.S. Energy Information Administration, 2 August 2021. [Online]. Available: https://www.eia.gov/energyexplained/use-of-energy/industry.php.

[39] P. Schober, C. Boer and L. A. Schwarte, "Correlation Coefficients: Appropriate Use and Interpretation," *Anesthesia and analgesia,* p. 1763–1768, 2018.

[40] D. E. Hinkle, W. Wiersma, and S. G. Jurs, Applied Statistics for the Behavioral Sciences 5th Edition, Boston: Houghton Mifflin, 2002.

[41] L. Cohen, P. Jarvis and J. Fowler, Practical Statistics for Field Biology 2nd Edition, Wiley, 1998.

[42] R. Taylor, "Interpretation of the Correlation Coefficient: A Basic Review," *Journal of Diagnostic Medical Sonography,* vol. 6, no. 1, pp. 35-39, 1990.

[43] "ISO-3166-1 alpha-2," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2.

[44] "Capacity factor," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Capacity_factor#cite_note-1.

APPENDIX A: THE HISTOGRAMS, THE SCATTER PLOTS, AND THE KERNEL DENSITY ESTIMATE (KDE) PLOTS BETWEEN WEPP DATASET AND MSCI (2018) DATASET

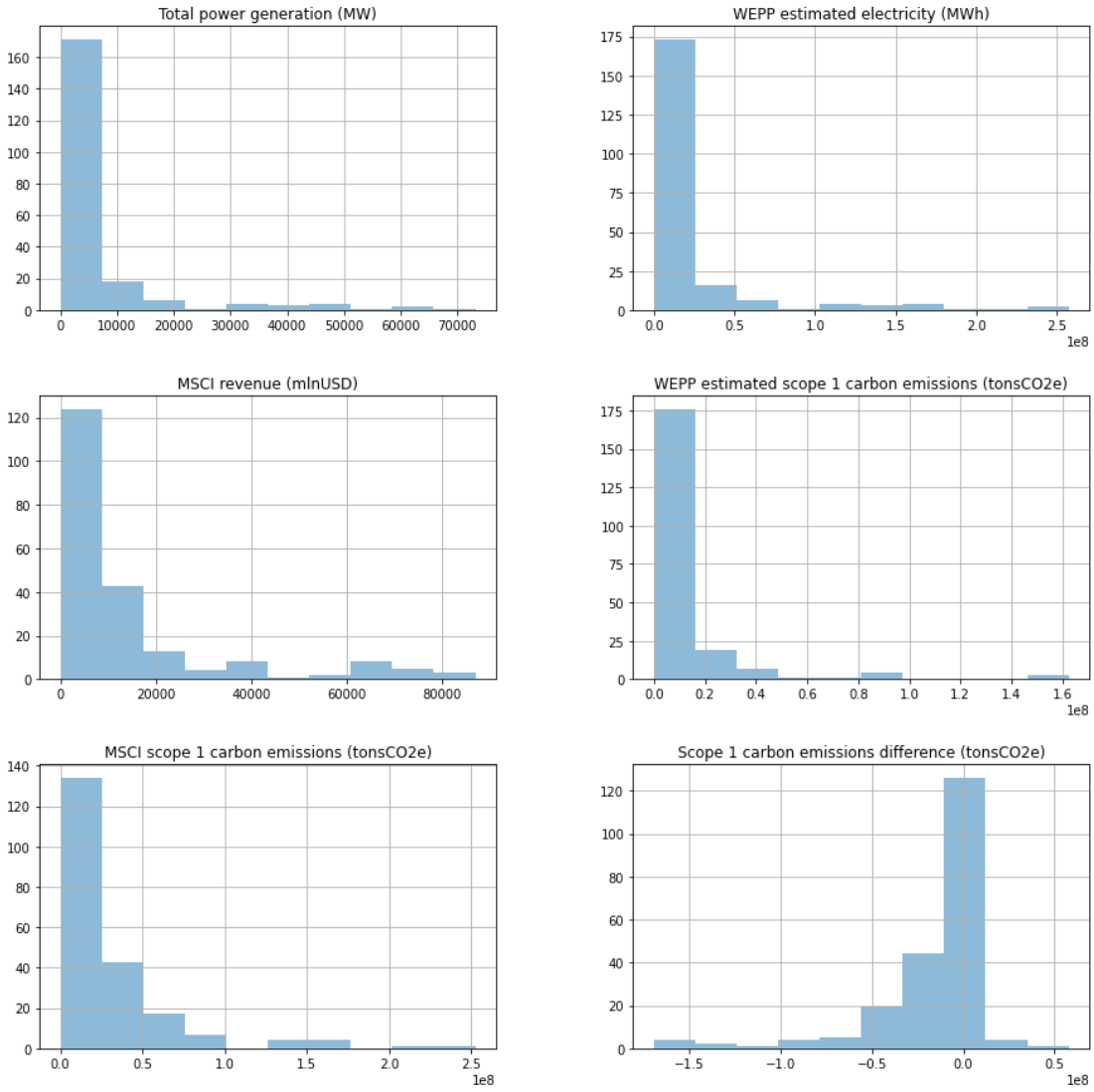

FIG. 13: The histograms of quantitative factors in WEPP dataset and MSCI (2018) dataset, enriched by Orbis through CorpIndex, i.e., total power generation (MW), WEPP estimated electricity (MWh), MSCI revenue (mln USD), WEPP estimated scope 1 carbon emissions (tonsCO_2e), MSCI scope 1 carbon emissions (tonsCO_2e), scope 1 carbon emissions difference (tonsCO_2e). The histogram of the scope 1 carbon emissions difference shows it is heavily right-skewed.

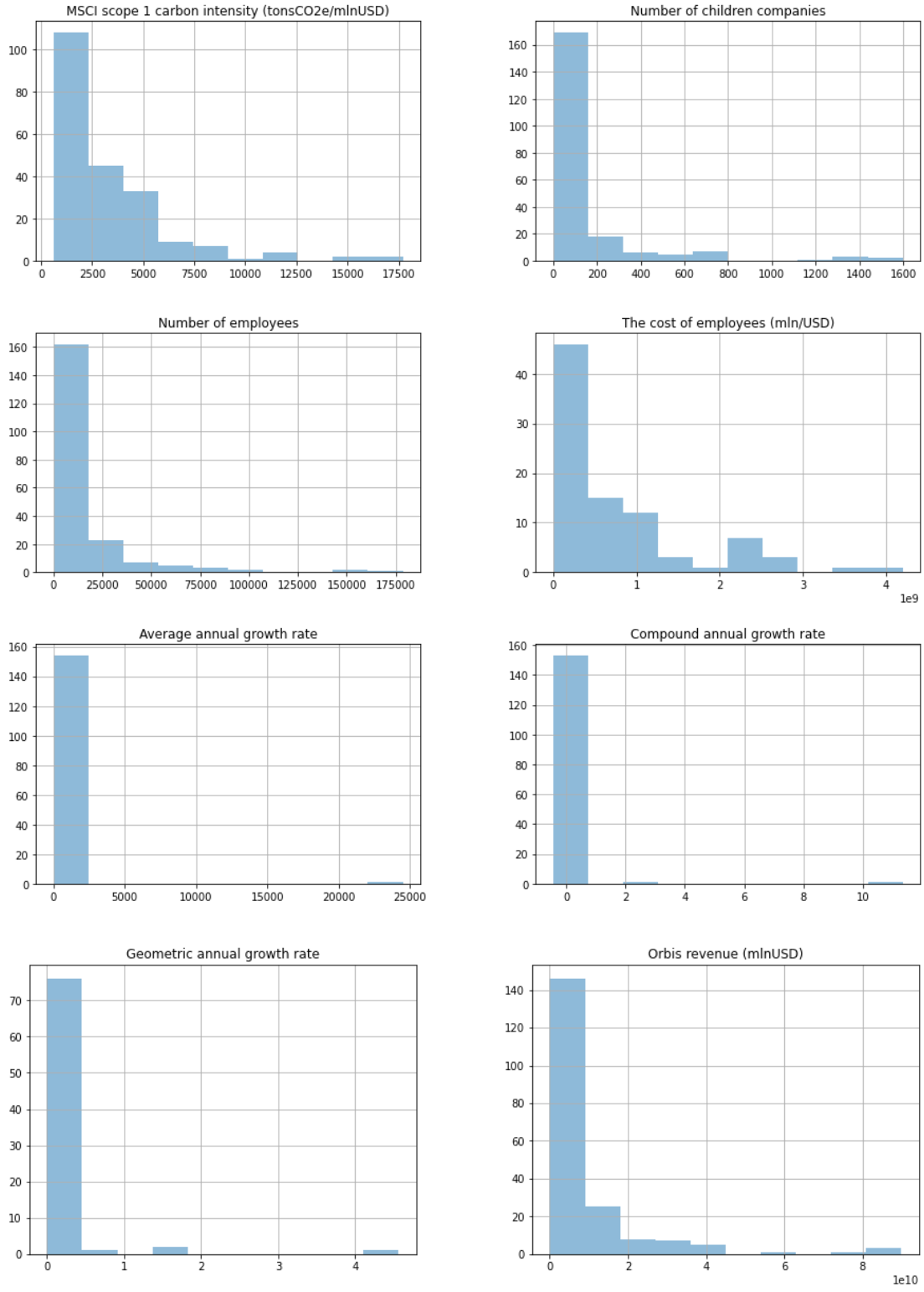

FIG. 14: The histograms of quantitative factors in WEPP dataset and MSCI (2018) dataset, enriched by Orbis through CorpIndex, i.e., MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), number of children companies, number of employees, the cost of employees (mln/USD), average annual growth rate, compound annual growth rate, geometric annual growth rate, and Orbis revenue (mln USD).

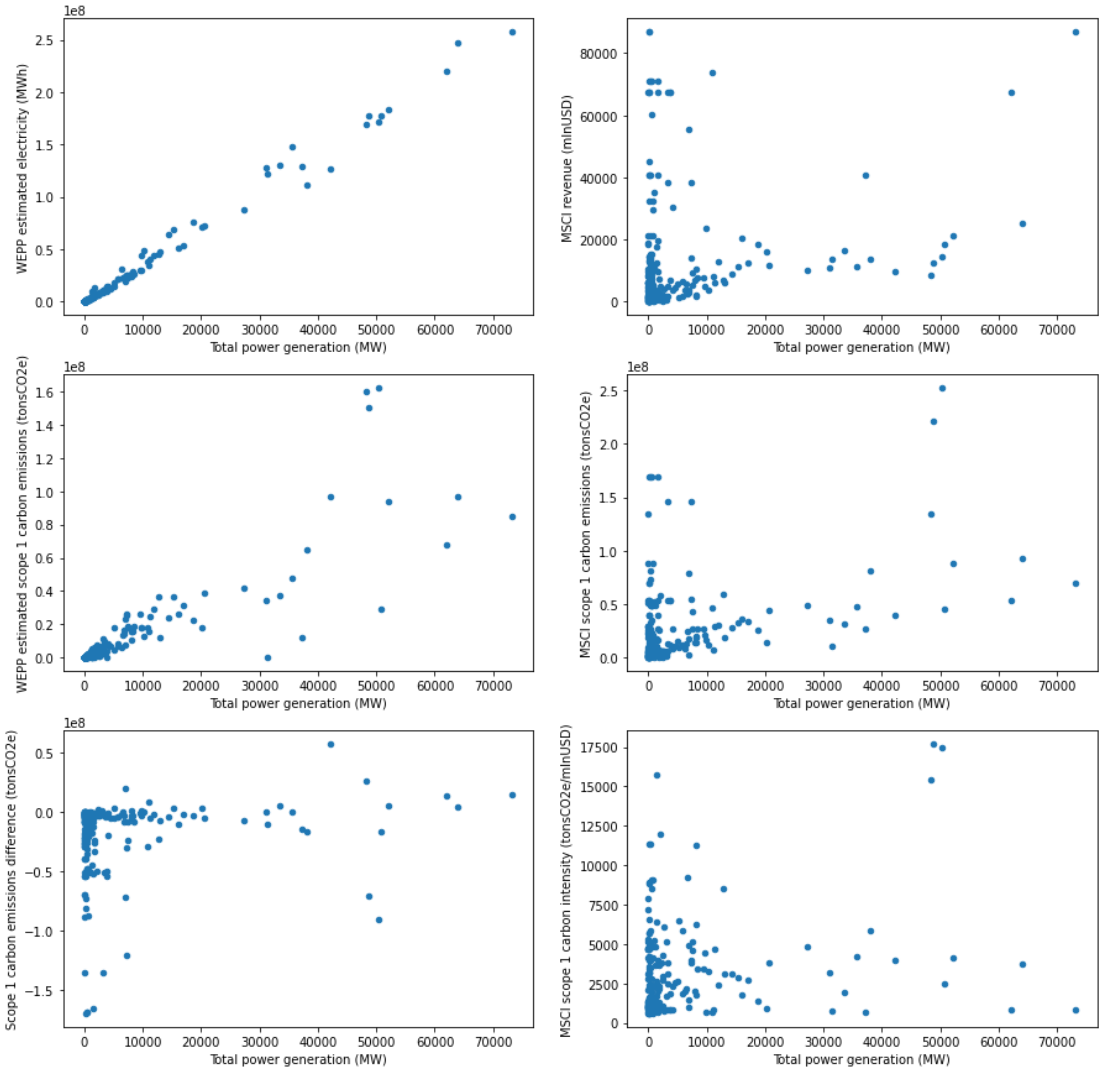

FIG. 15: The scatter plots between total power generation (MW) and other quantitative factors in WEPP dataset and MSCI (2018) dataset, enriched by Orbis through CorpIndex, i.e., WEPP estimated electricity (MWh), MSCI revenue (mln USD), WEPP estimated scope 1 carbon emissions (tonsCO_2e), MSCI scope 1 carbon emissions (tonsCO_2e), scope 1 carbon emissions difference (tonsCO_2e), and MSCI scope 1 carbon intensity (tonsCO_2e/mln USD).

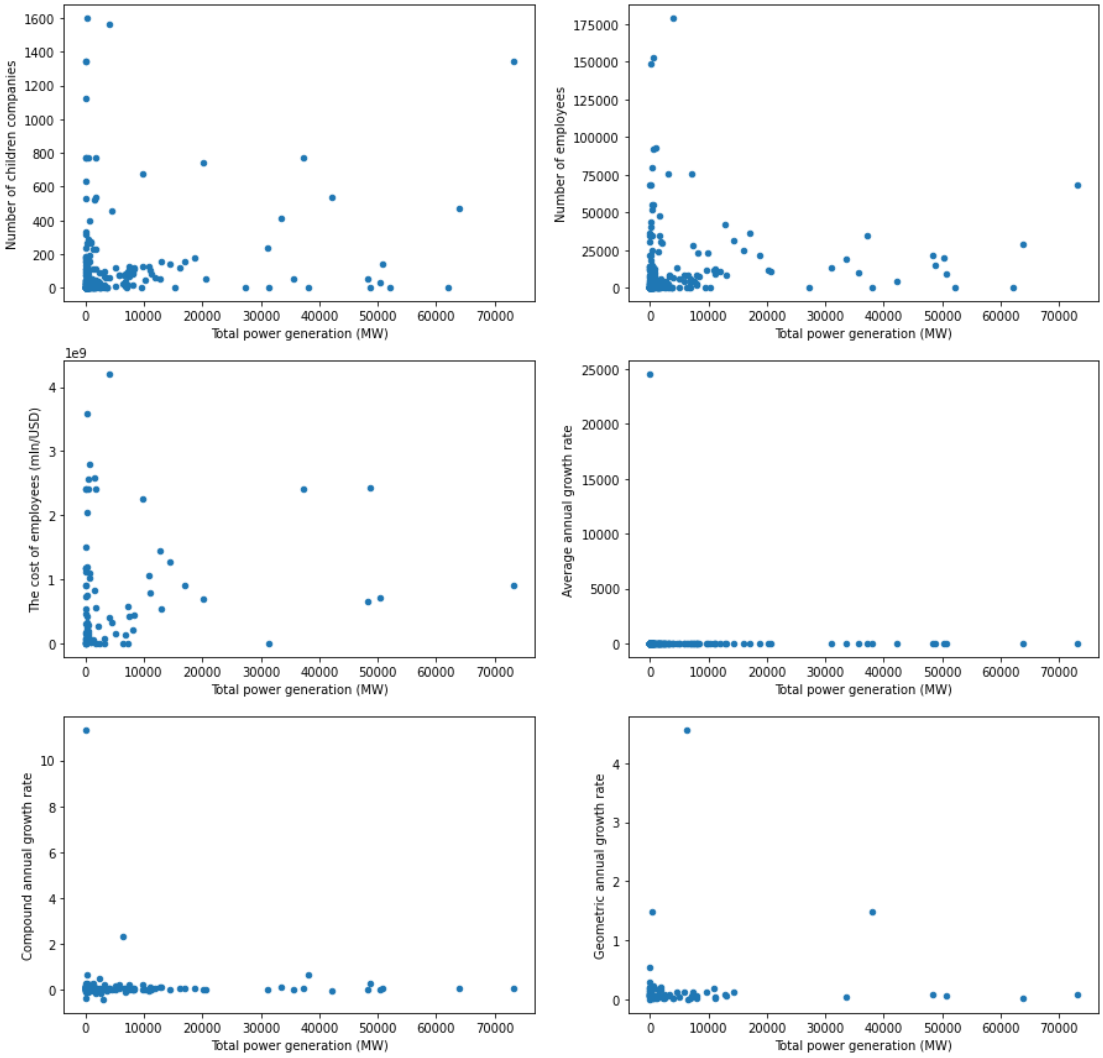

FIG. 16: The scatter plots between total power generation (MW) and other quantitative factors in WEPP dataset and MSCI (2018) dataset, enriched by Orbis through CorpIndex, i.e., number of children companies, number of employees, the cost of employees (mln/USD), average annual growth rate, compound annual growth rate, and geometric annual growth rate.

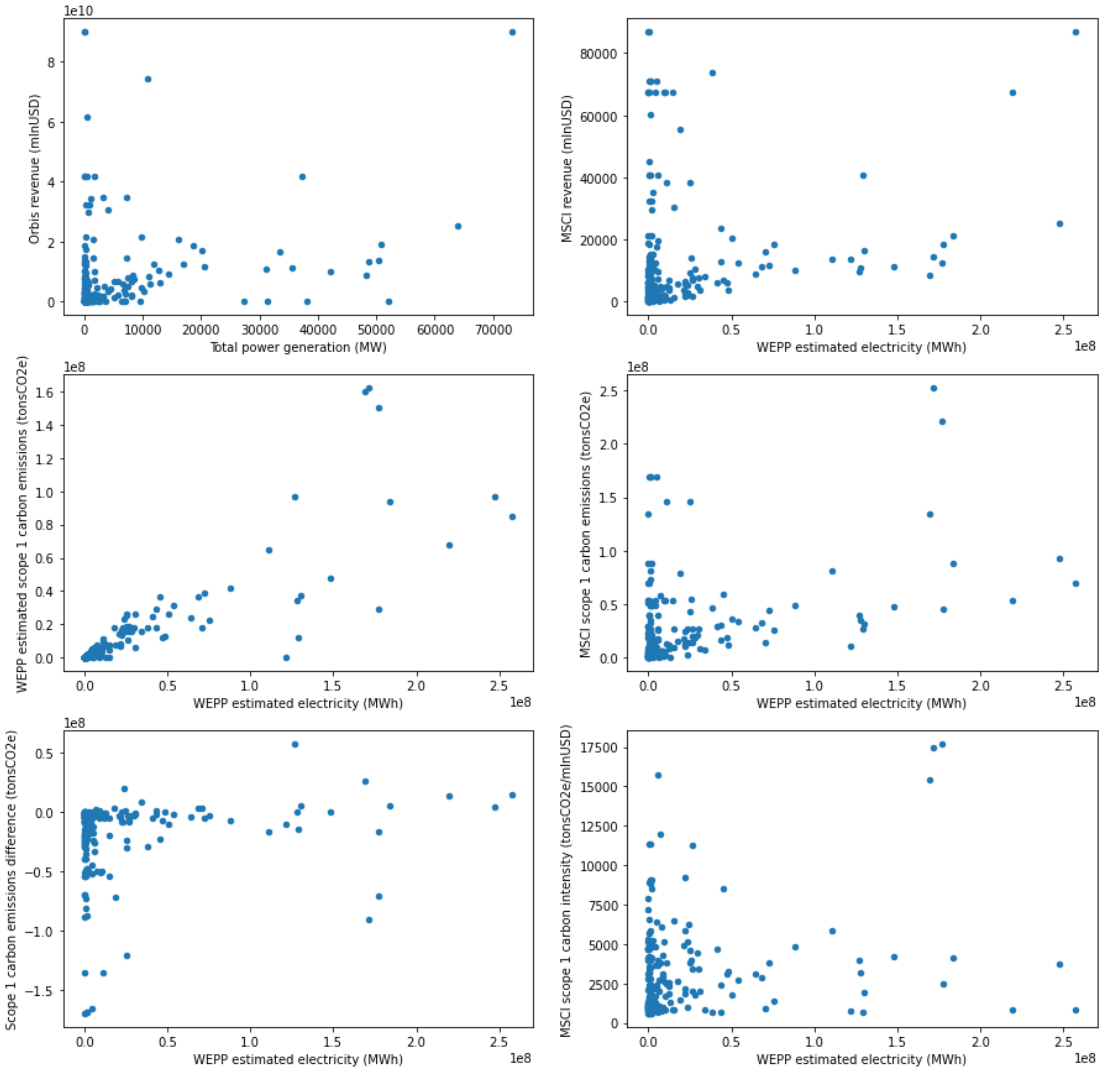

FIG. 17: The scatter plot between total power generation (MW) and Orbis revenue (mln USD). And the scatter plots between WEPP estimated electricity (MWh) and other quantitative factors, i.e., MSCI revenue (mln USD), WEPP estimated scope 1 carbon emissions (tonsCO_2e), MSCI scope 1 carbon emissions (tonsCO_2e), scope 1 carbon emissions difference (tonsCO_2e), and MSCI scope 1 carbon intensity (tonsCO_2e/mln USD).

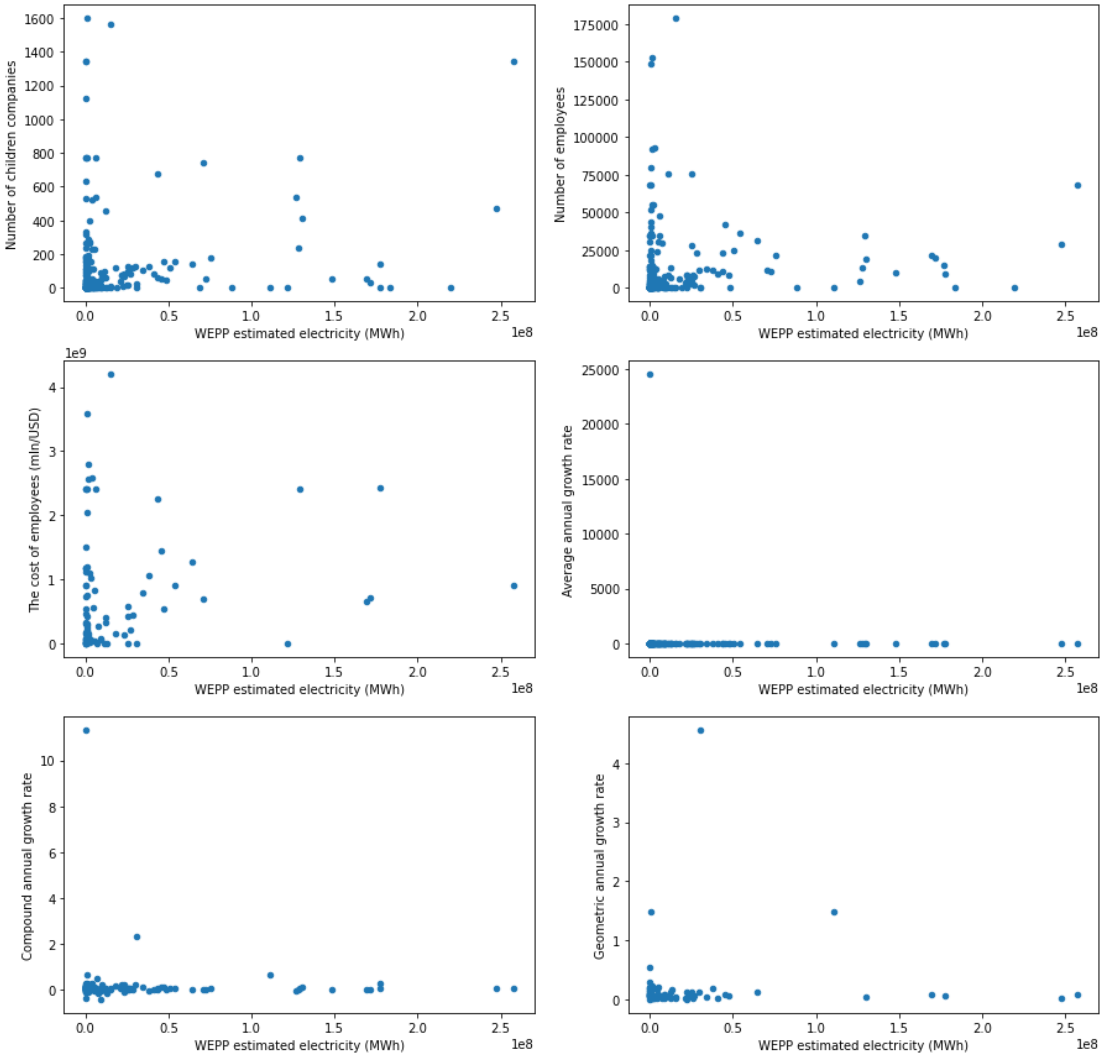

FIG. 18: The scatter plot between total power generation (MW) and Orbis revenue (mln USD). And the scatter plots between WEPP estimated electricity (MWh) and other quantitative factors, i.e., MSCI revenue (mln USD), WEPP estimated scope 1 carbon emissions (tonsCO_2e), MSCI scope 1 carbon emissions (tonsCO_2e), scope 1 carbon emissions difference (tonsCO_2e), and MSCI scope 1 carbon intensity (tonsCO_2e/mln USD).

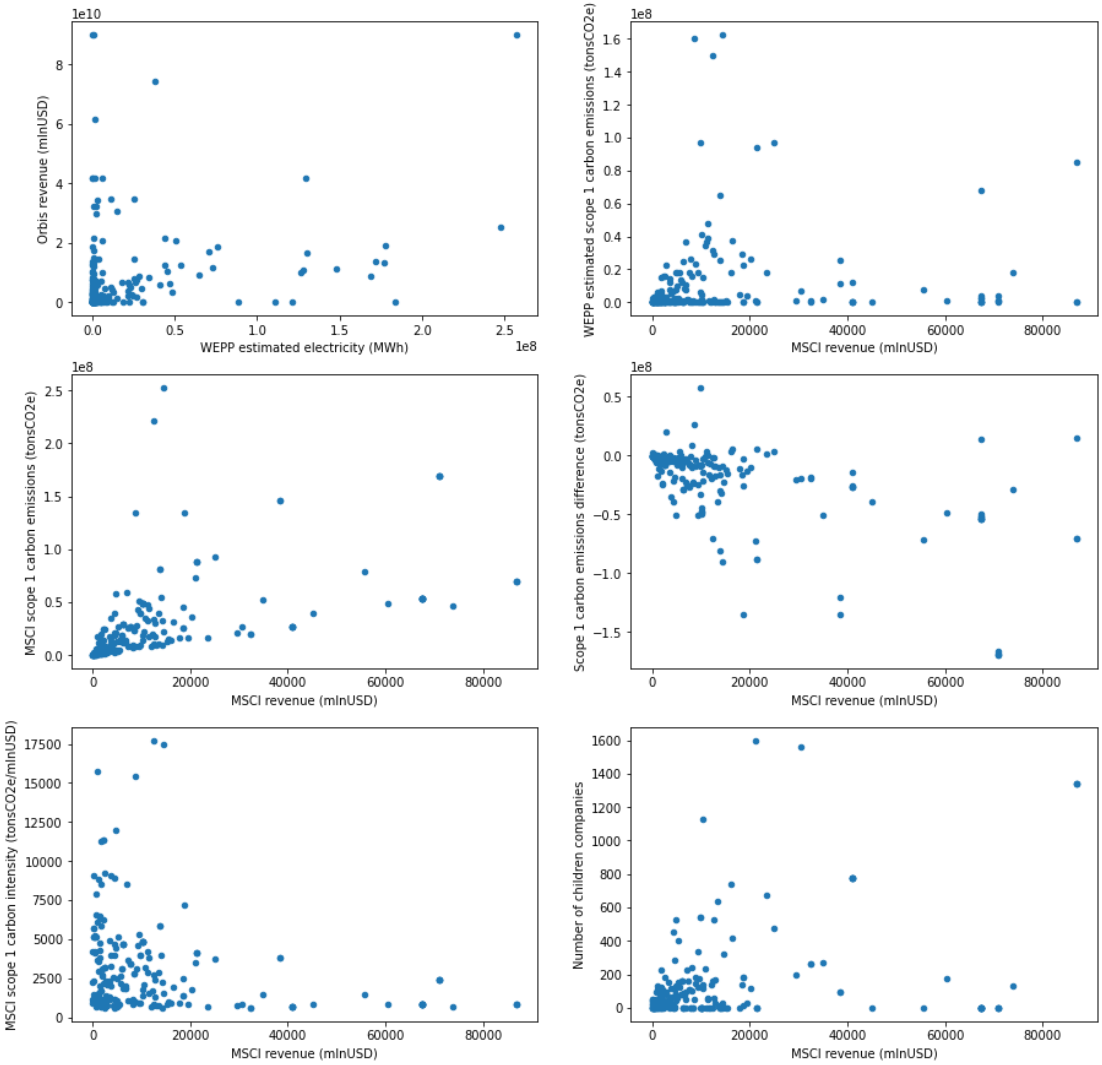

FIG. 19: The scatter plot between WEPP estimated electricity (MWh) and Orbis revenue (mln USD). And the scatter plots between MSCI revenue (mln USD) and other quantitative factors, i.e., WEPP estimated scope 1 carbon emissions (tonsCO_2e), MSCI scope 1 carbon emissions (tonsCO_2e), scope 1 carbon emissions difference (tonsCO_2e), MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), and number of children companies.

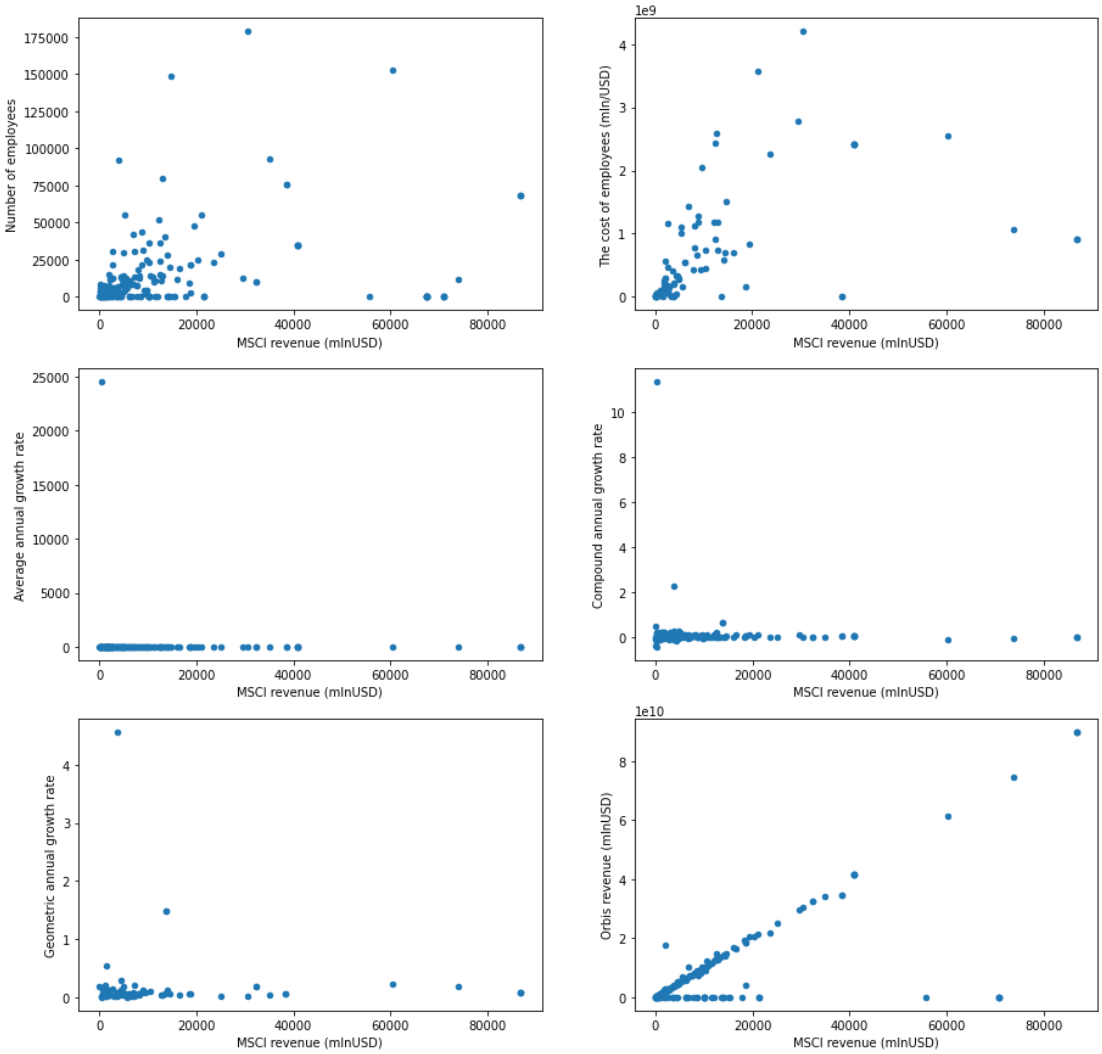

FIG. 20: The scatter plots between MSCI revenue (mln USD) and other quantitative factors, i.e., number of employees, the cost of employees (mln/USD), average annual growth rate, compound annual growth rate, geometric annual growth rate, and Orbis revenue (mln USD).

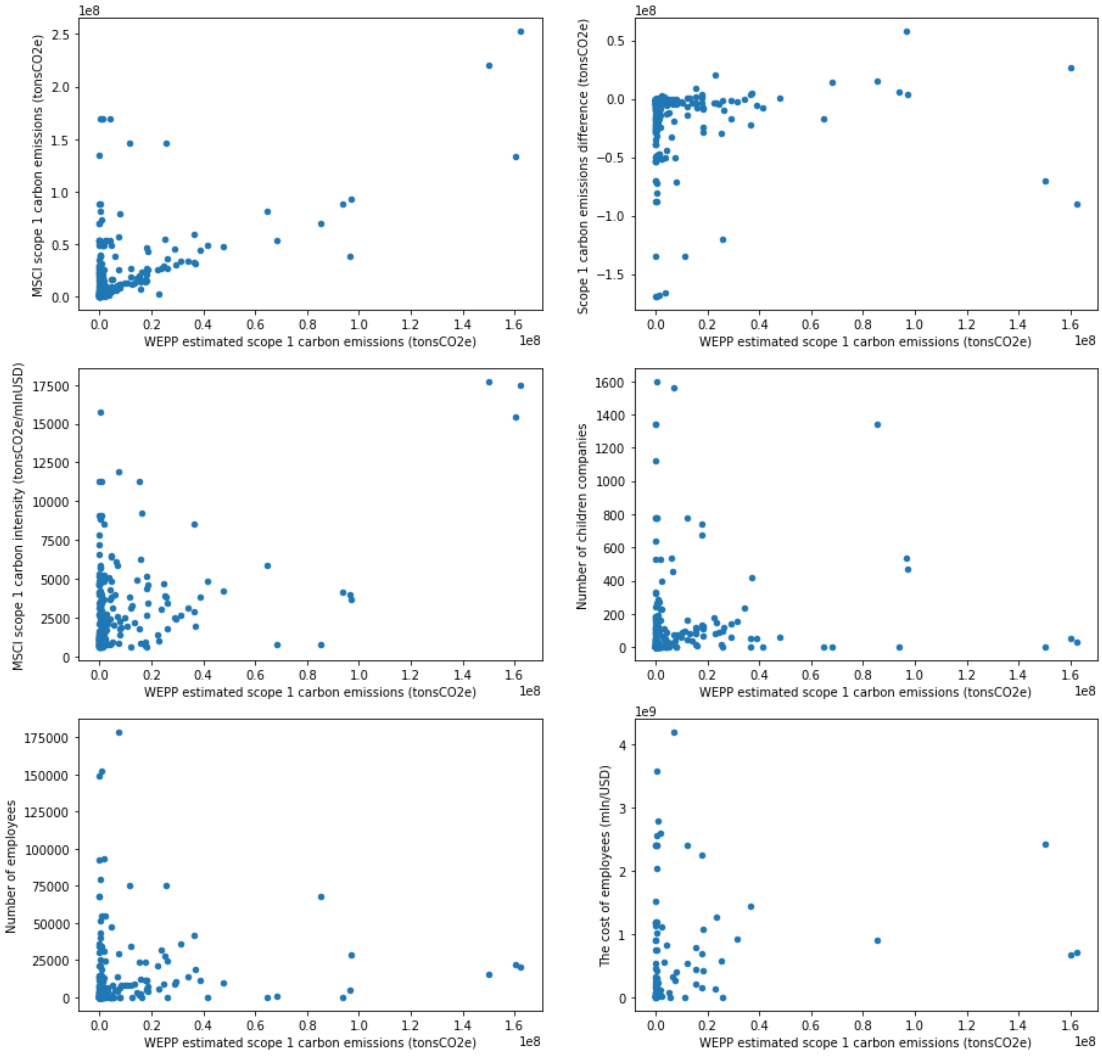

FIG. 21: The scatter plots between WEPP estimated scope 1 carbon emissions (tonsCO_2e) and other quantitative factors, i.e., MSCI scope 1 carbon emissions (tonsCO_2e), scope 1 carbon emissions difference (tonsCO_2e), MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), number of children companies, number of employees, and the cost of employees (mln/USD).

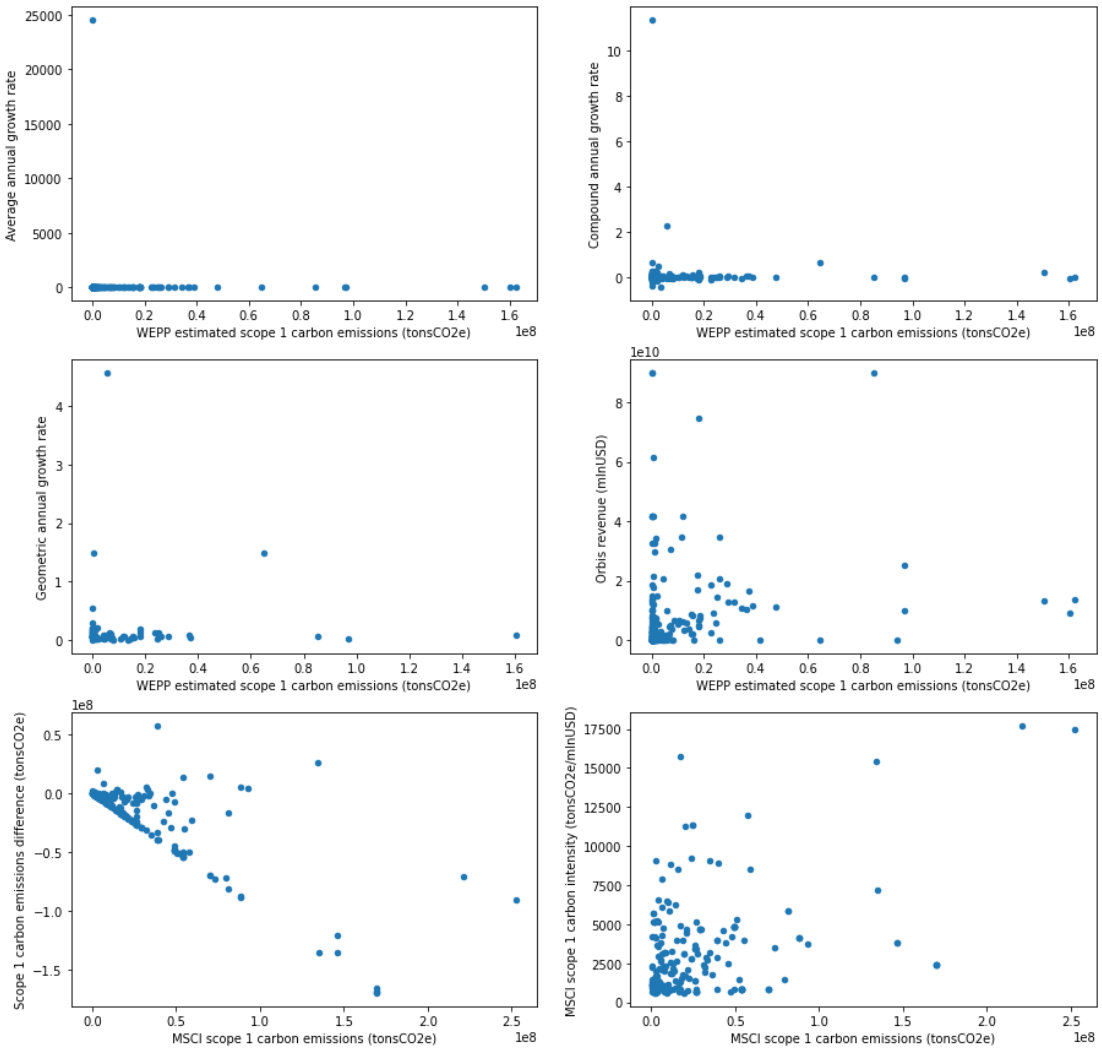

FIG. 22: The scatter plots between WEPP estimated scope 1 carbon emissions (tonsCO_2e) and other quantitative factors, i.e., average annual growth rate, compound annual growth rate, geometric annual growth rate, and Orbis revenue (mln USD). And the scatter plots between MSCI scope 1 carbon emissions (tonsCO_2e) and other quantitative factors, i.e., scope 1 carbon emissions difference (tonsCO_2e), and MSCI scope 1 carbon intensity (tonsCO_2e/mln USD).

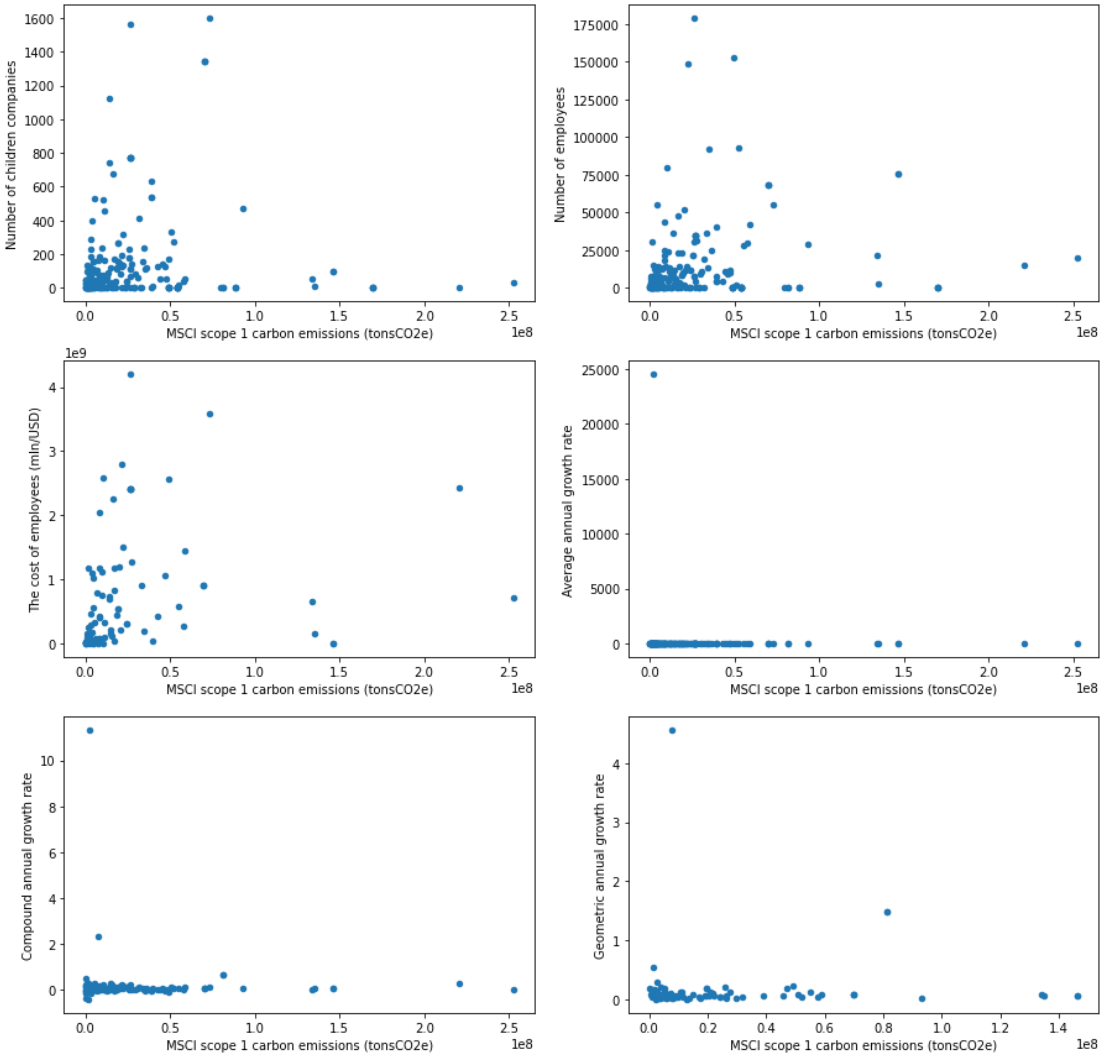

FIG. 23: The scatter plots between MSCI scope 1 carbon emissions (tonsCO_2e) and other quantitative factors, i.e., number of children companies, number of employees, the cost of employees (mln/USD), average annual growth rate, compound annual growth rate, and geometric annual growth rate.

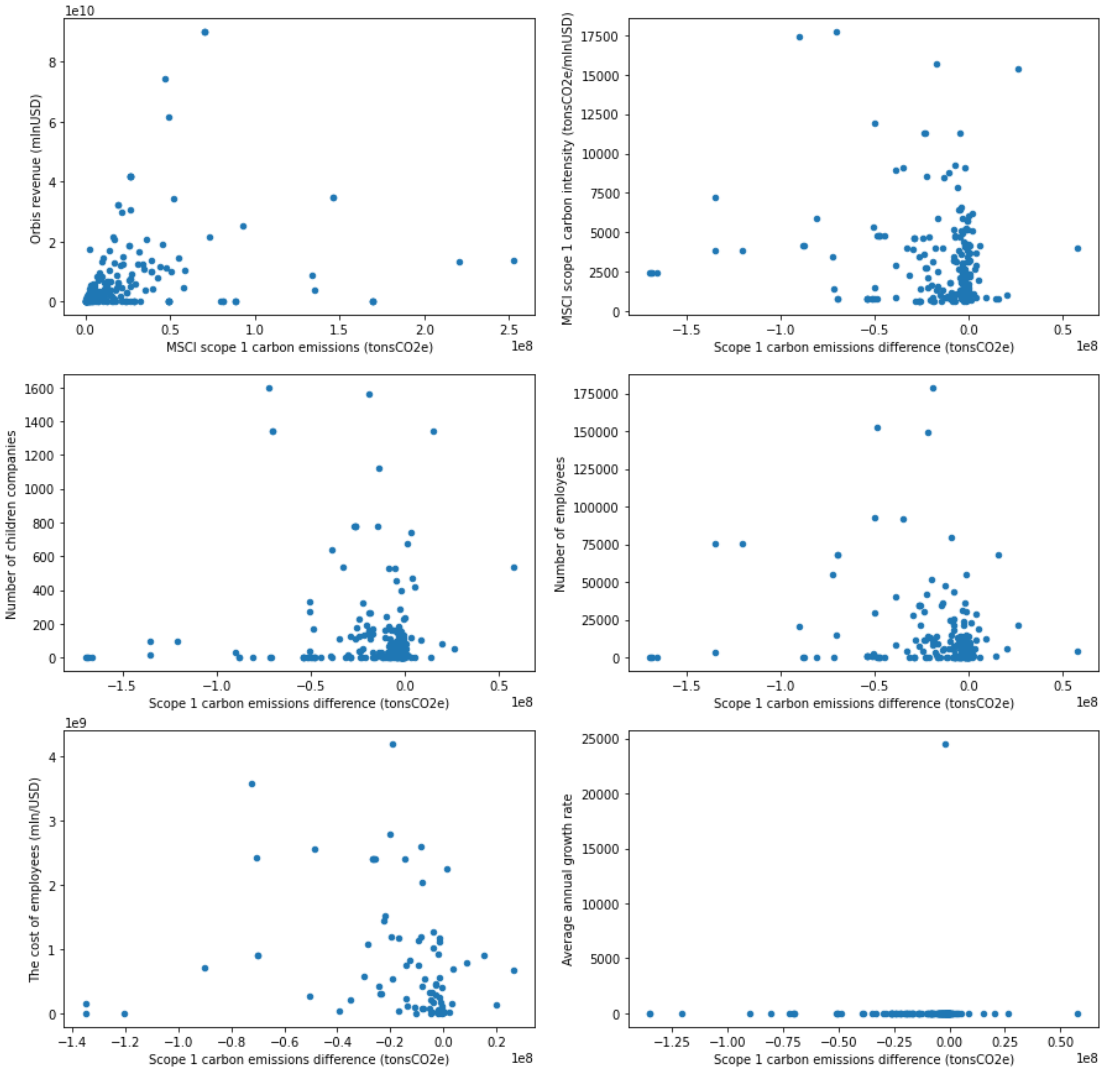

FIG. 24: The scatter plot between MSCI scope 1 carbon emissions (tonsCO_2e) and Orbis revenue (mln USD). And the scatter plots between scope 1 carbon emissions difference (tonsCO_2e) and other quantitative factors, i.e., MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), number of children companies, number of employees, the cost of employees (mln/USD), and average annual growth rate.

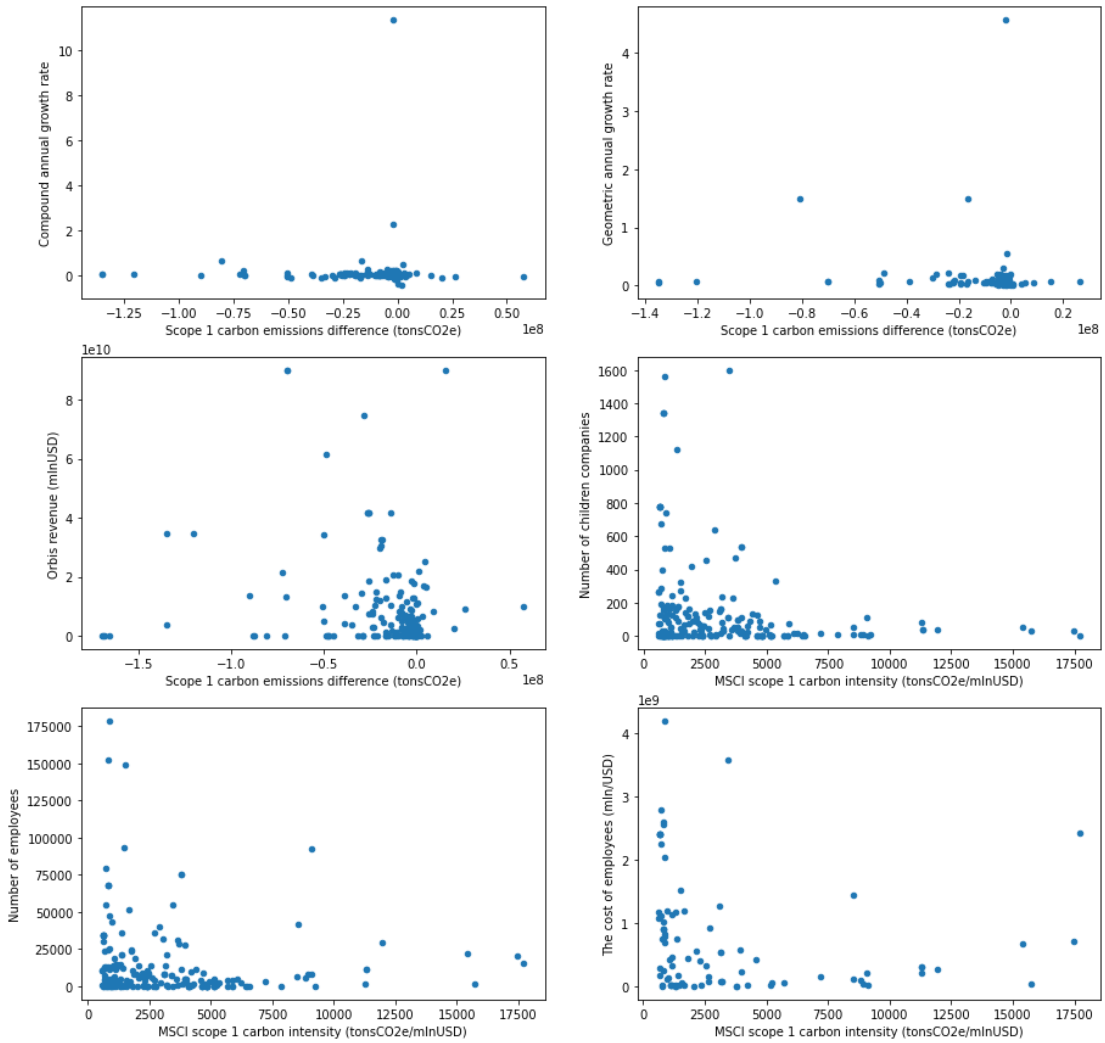

FIG. 25: The scatter plots between scope 1 carbon emissions difference (tonsCO_2e) and other quantitative factors, i.e., compound annual growth rate, geometric annual growth rate, and Orbis revenue (mln USD). And the scatter plots between MSCI scope 1 carbon intensity (tonsCO_2e/mln USD) and other quantitative factors, i.e., number of children companies, number of employees, and the cost of employees (mln/USD).

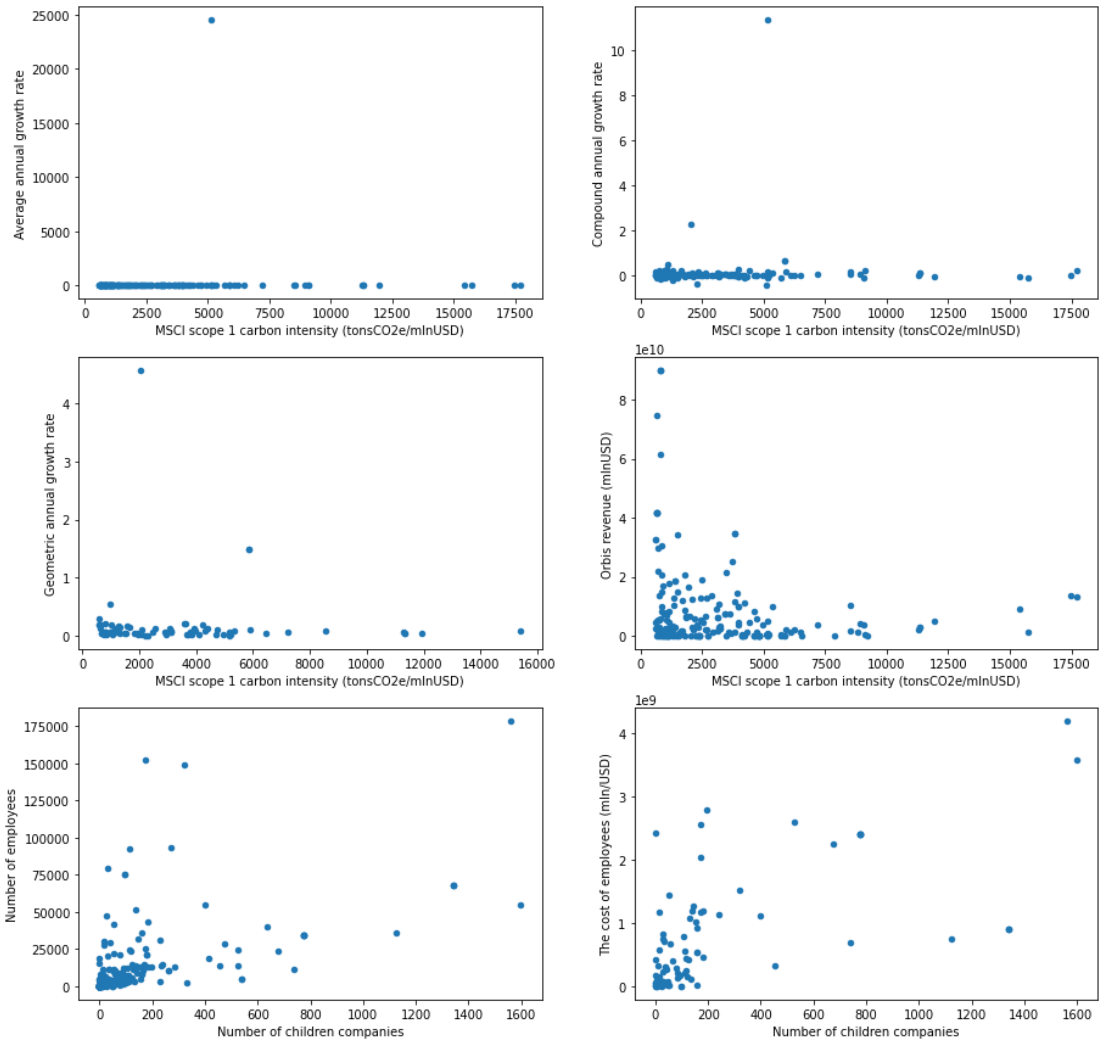

FIG. 26: The scatter plots between MSCI scope 1 carbon intensity (tonsCO_2e/mln USD) and other quantitative factors, i.e., average annual growth rate, compound annual growth rate, geometric annual growth rate, and Orbis revenue (mln USD). And the scatter plots between number of children companies and other quantitative factors, i.e., number of employees, and the cost of employees (mln/USD).

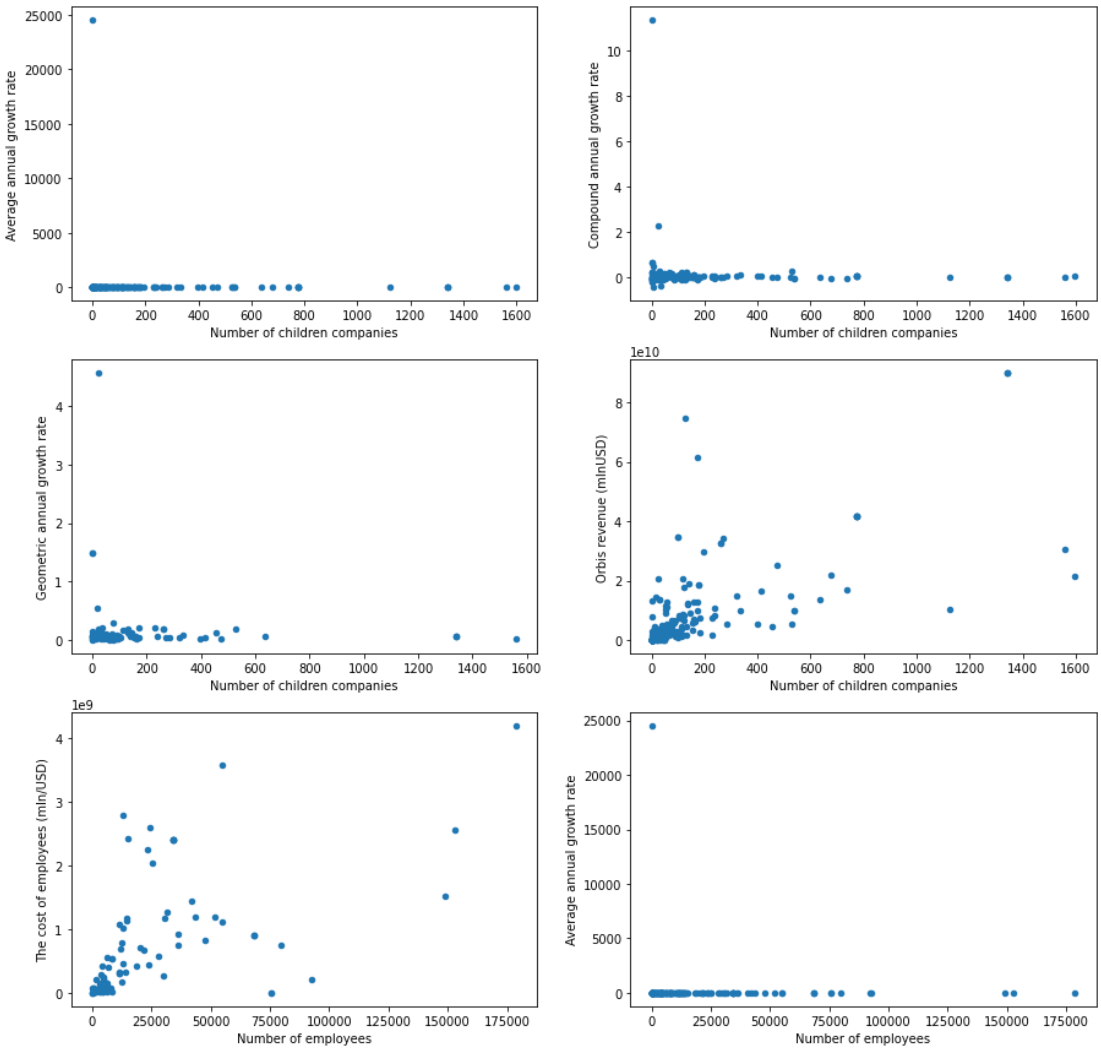

FIG. 27: The scatter plots between number of children companies and other quantitative factors, i.e., average annual growth rate, compound annual growth rate, geometric annual growth rate, and Orbis revenue (mln USD). And the scatter plots between number of employees and other quantitative factors, i.e., the cost of employees (mln/USD), and average annual growth rate.

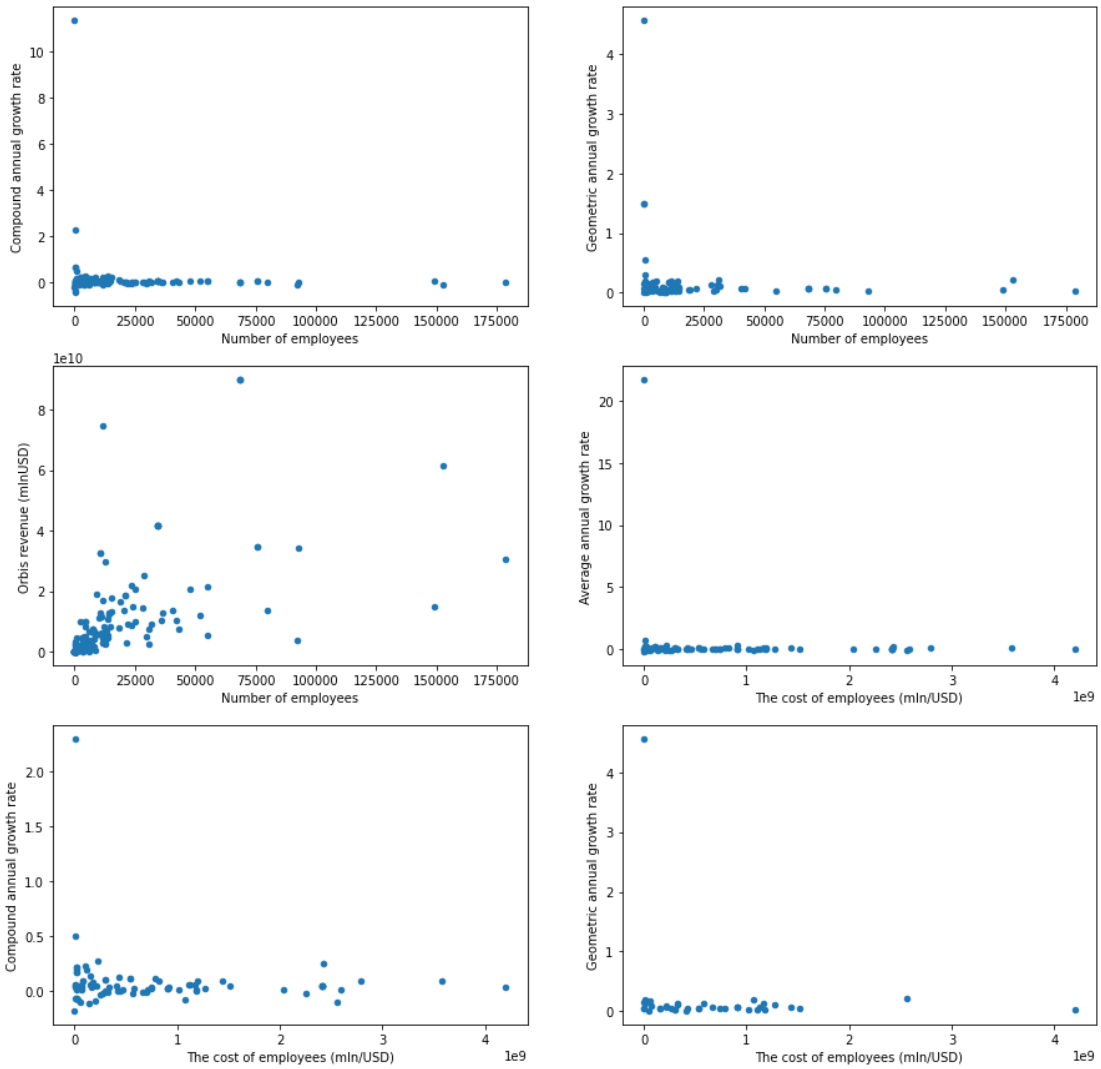

FIG. 28: The scatter plots between number of employees and other quantitative factors, i.e., compound annual growth rate, and geometric annual growth rate, and Orbis revenue (mln USD). And the scatter plots between the cost of employees (mln/USD) and other quantitative factors, i.e., average annual growth rate, compound annual growth rate, and geometric annual growth rate.

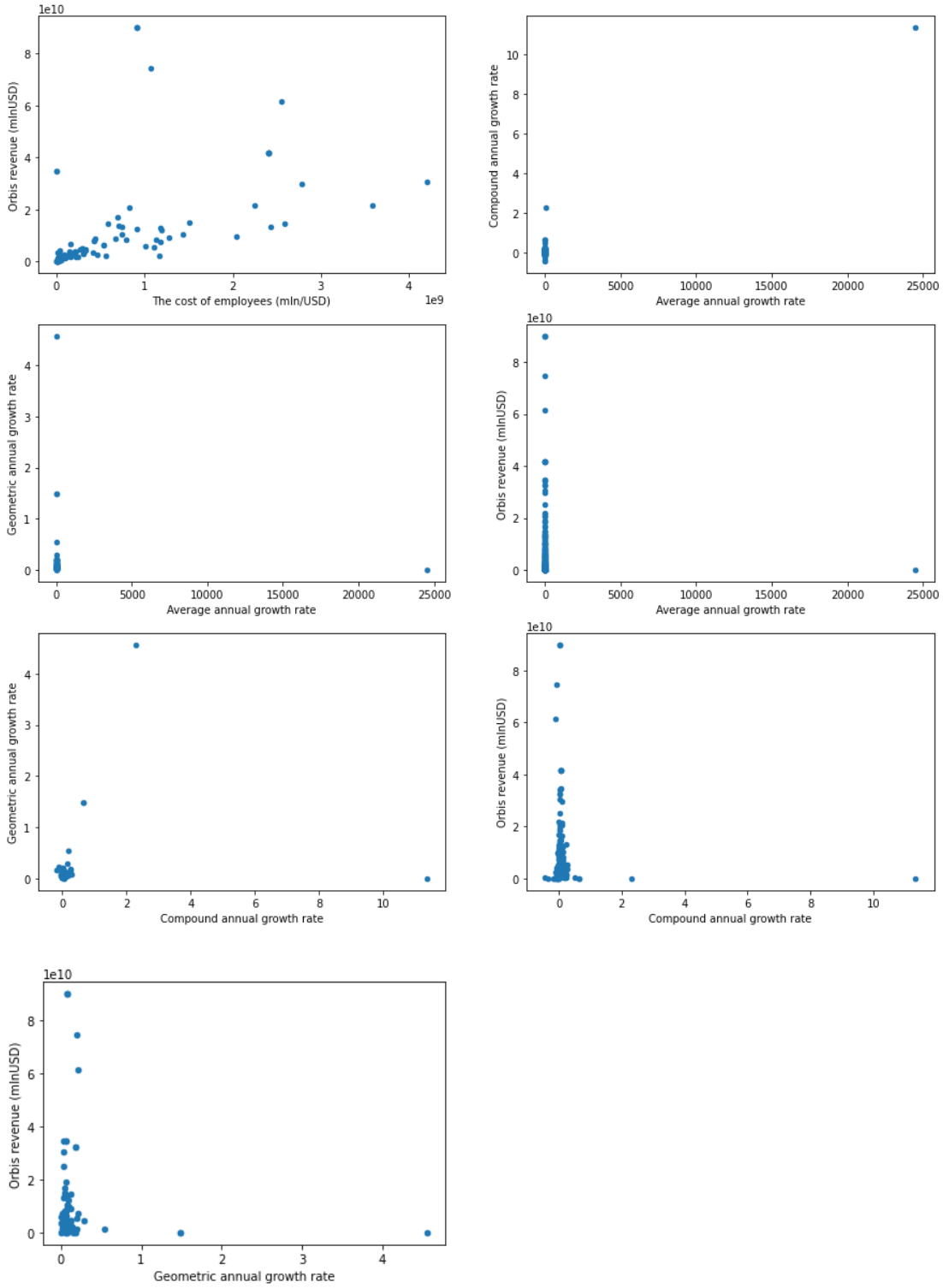

FIG. 29: The scatter plot between number of employees and Orbis revenue (mln USD). The scatter plots between average annual growth rate and other quantitative factors, i.e., compound annual growth rate, geometric annual growth rate, and Orbis revenue (mln USD). The scatter plots between compound annual growth rate and other quantitative factors, i.e., geometric annual growth rate, and Orbis revenue (mln USD). The scatter plot between geometric annual growth rate and Orbis revenue (mln USD).

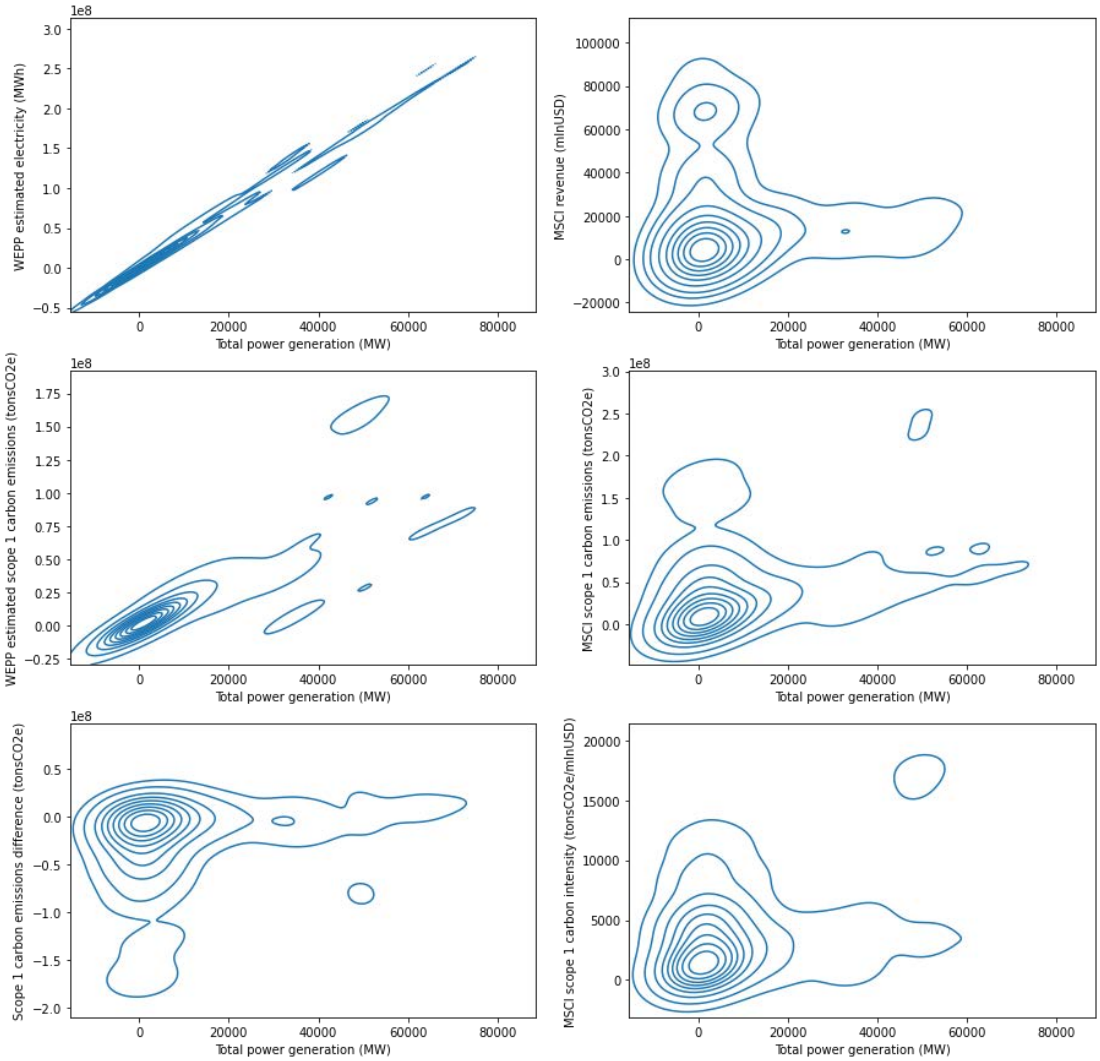

FIG. 30: The kernel density estimate (KDE) plots between total power generation (MW) and other quantitative factors in WEPP dataset and MSCI (2018) dataset, enriched by Orbis through CorpIndex, i.e., WEPP estimated electricity (MWh), MSCI revenue (mln USD), WEPP estimated scope 1 carbon emissions (tonsCO_2e), MSCI scope 1 carbon emissions (tonsCO_2e), scope 1 carbon emissions difference (tonsCO_2e), and MSCI scope 1 carbon intensity (tonsCO_2e/mln USD).

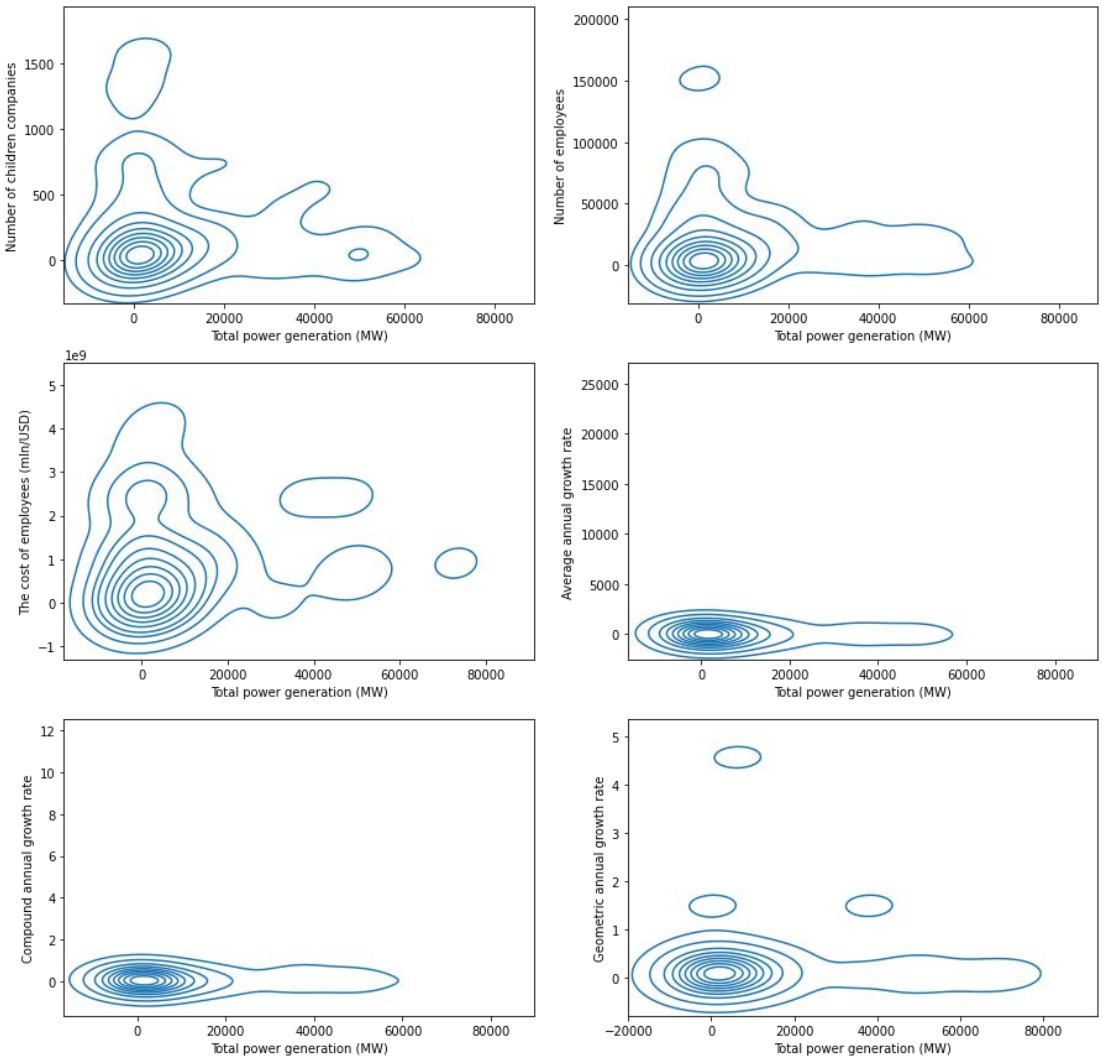

FIG. 31: The kernel density estimate (KDE) plots between total power generation (MW) and other quantitative factors in WEPP dataset and MSCI (2018) dataset, enriched by Orbis through CorpIndex, i.e., number of children companies, number of employees, the cost of employees (mln/USD), average annual growth rate, compound annual growth rate, and geometric annual growth rate.

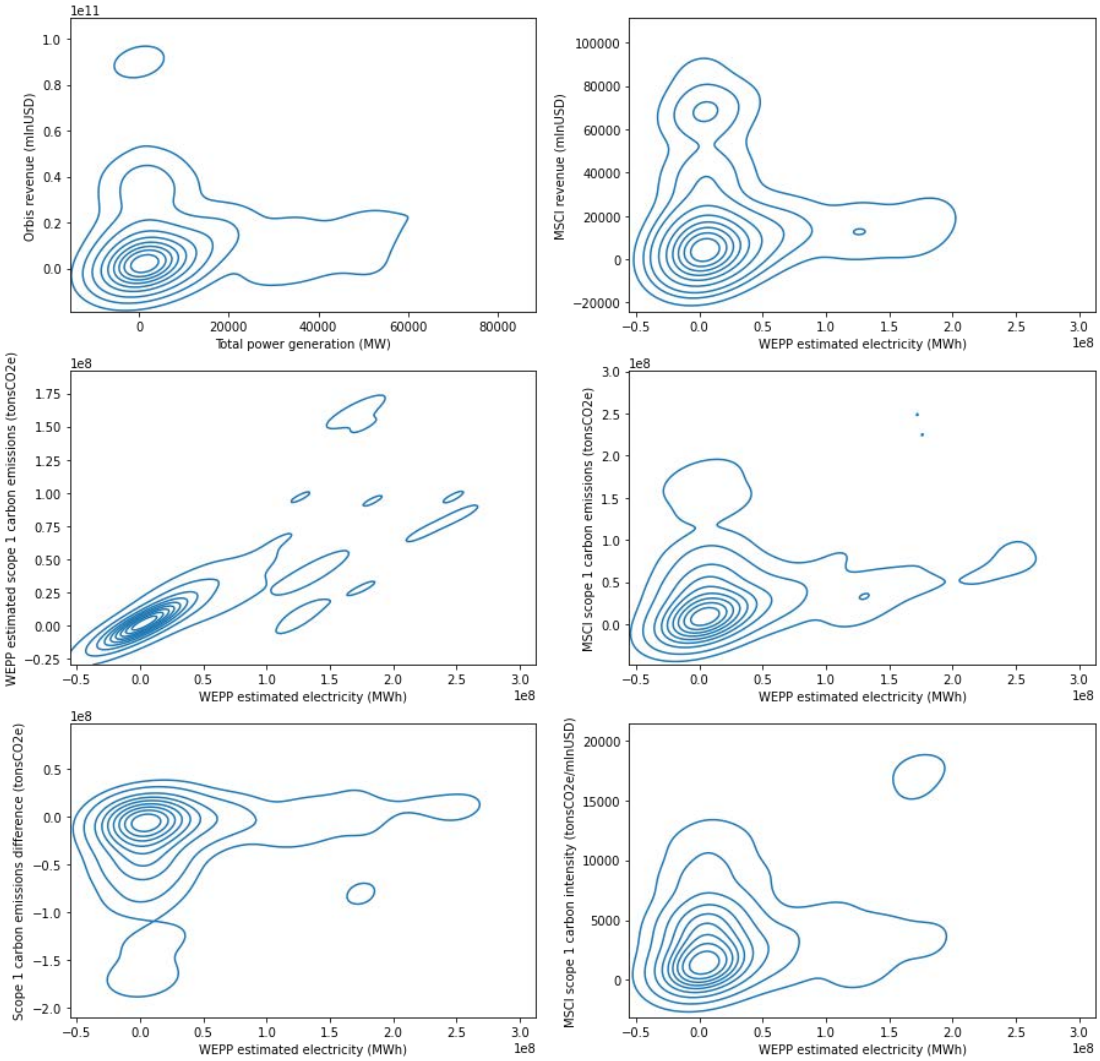

FIG. 32: The kernel density estimate (KDE) plot between total power generation (MW) and Orbis revenue (mln USD). And the kernel density estimate (KDE) plots between WEPP estimated electricity (MWh) and other quantitative factors, i.e., MSCI revenue (mln USD), WEPP estimated scope 1 carbon emissions (tonsCO_2e), MSCI scope 1 carbon emissions (tonsCO_2e), scope 1 carbon emissions difference (tonsCO_2e), and MSCI scope 1 carbon intensity (tonsCO_2e/mln USD).

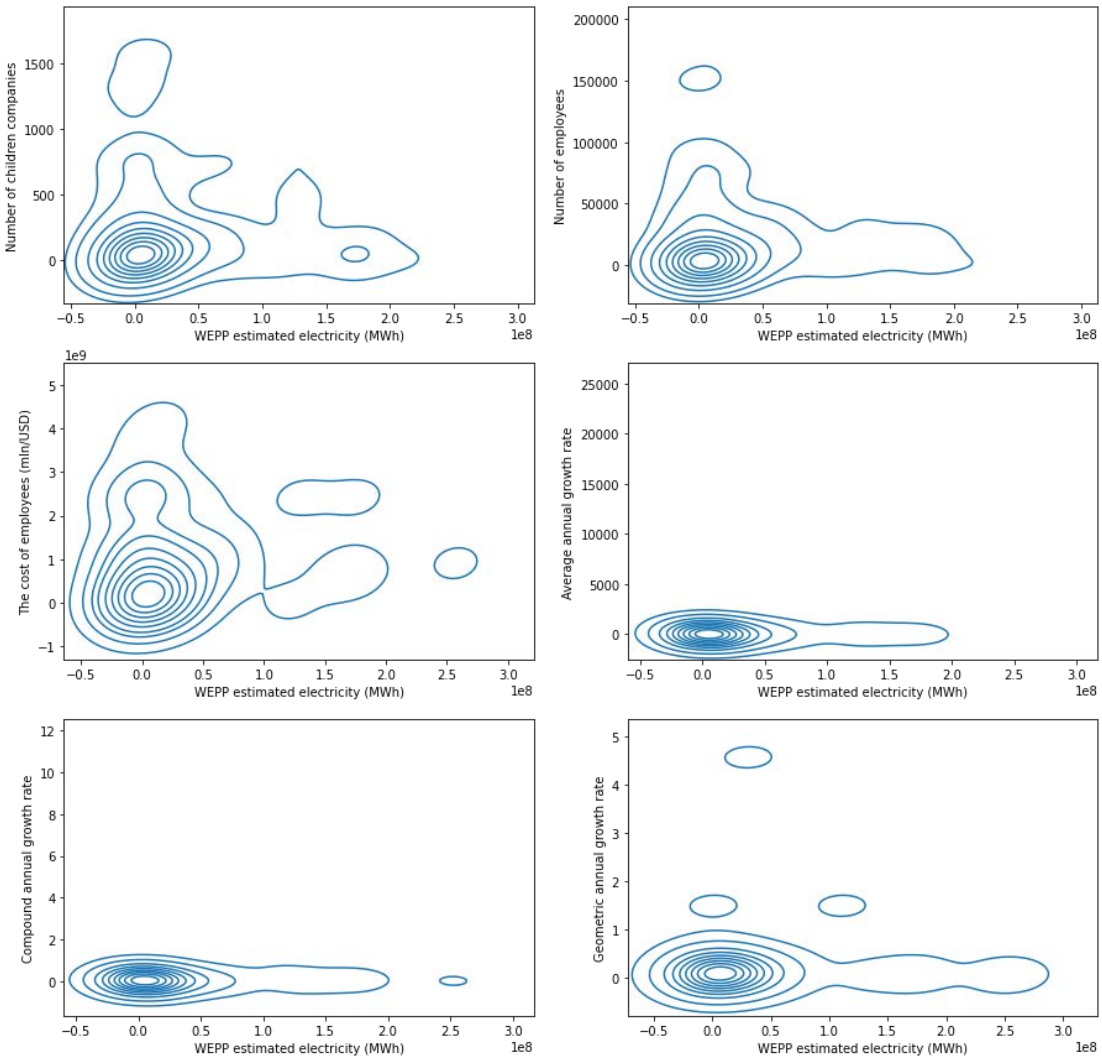

FIG. 33: The kernel density estimate (KDE) plot between total power generation (MW) and Orbis revenue (mln USD). And the kernel density estimate (KDE) plots between WEPP estimated electricity (MWh) and other quantitative factors, i.e., MSCI revenue (mln USD), WEPP estimated scope 1 carbon emissions (tonsCO_2e), MSCI scope 1 carbon emissions (tonsCO_2e), scope 1 carbon emissions difference (tonsCO_2e), and MSCI scope 1 carbon intensity (tonsCO_2e/mln USD).

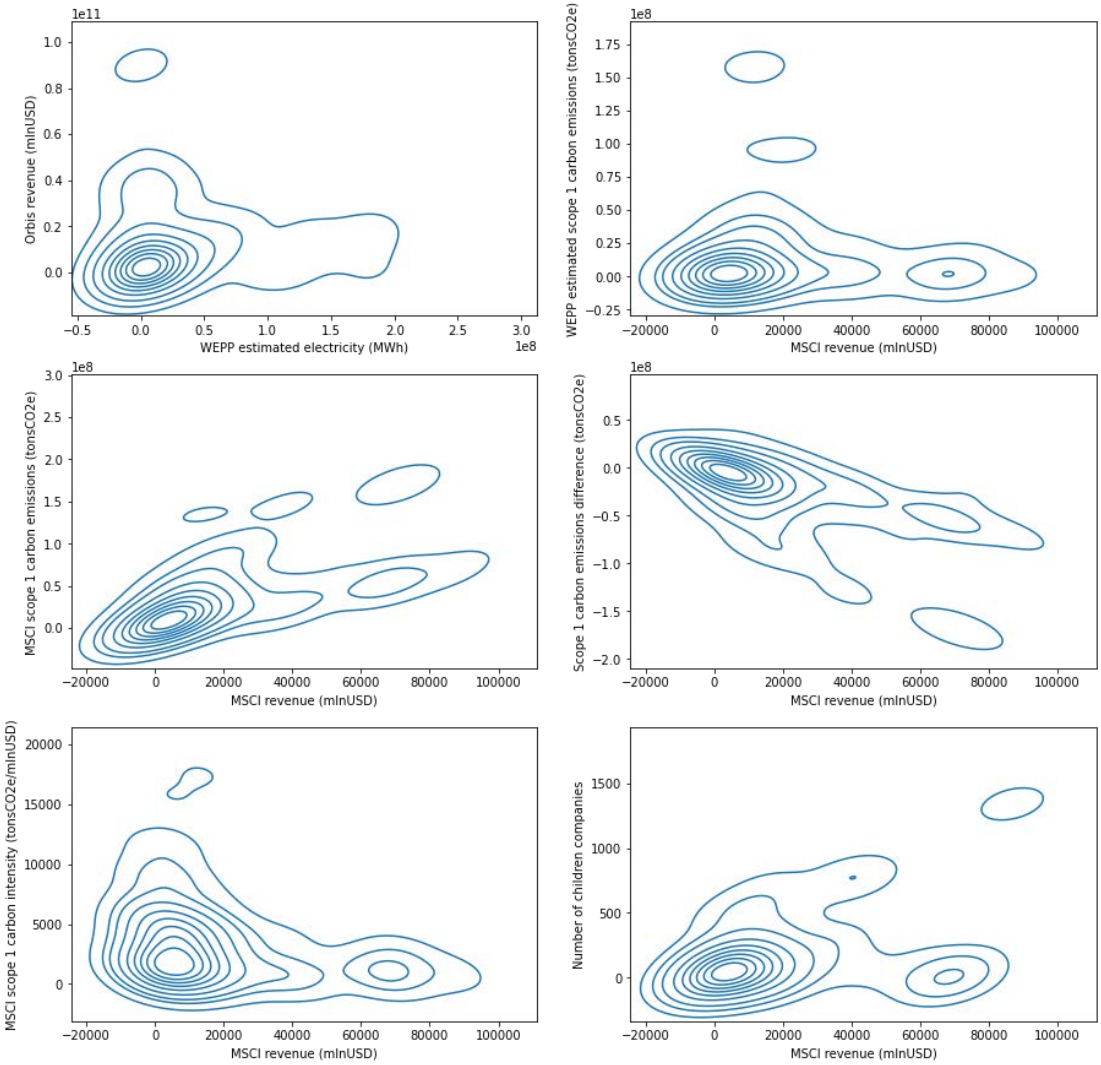

FIG. 34: The kernel density estimate (KDE) plot WEPP estimated electricity (MWh) and Orbis revenue (mln USD). And the kernel density estimate (KDE) plots between MSCI revenue (mln USD) and other quantitative factors, i.e., WEPP estimated scope 1 carbon emissions (tonsCO_2e), MSCI scope 1 carbon emissions (tonsCO_2e), scope 1 carbon emissions difference (tonsCO_2e), MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), and number of children companies.

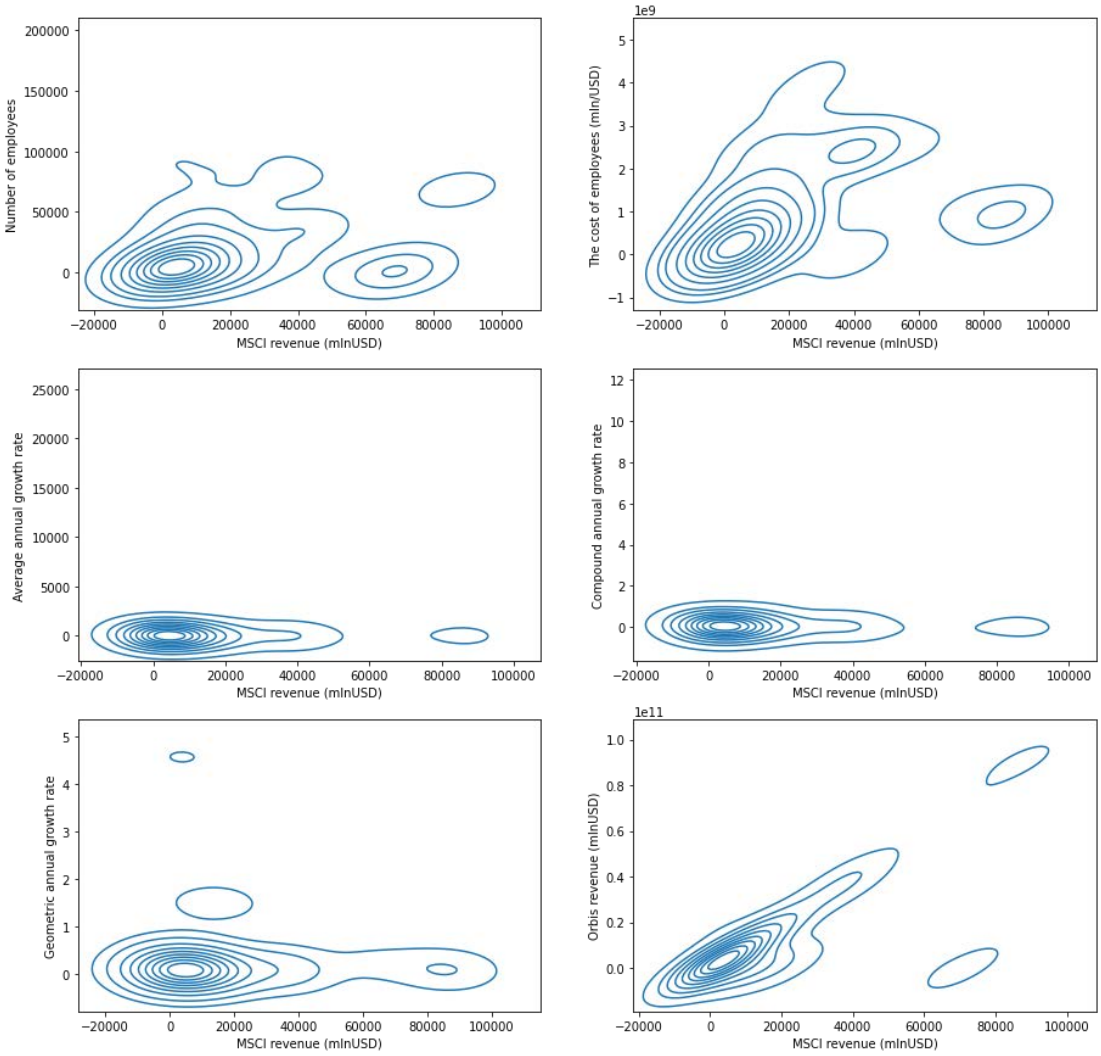

FIG. 35: The kernel density estimate (KDE) plots between MSCI revenue (mln USD) and other quantitative factors, i.e., number of employees, the cost of employees (mln/USD), average annual growth rate, compound annual growth rate, geometric annual growth rate, and Orbis revenue (mln USD).

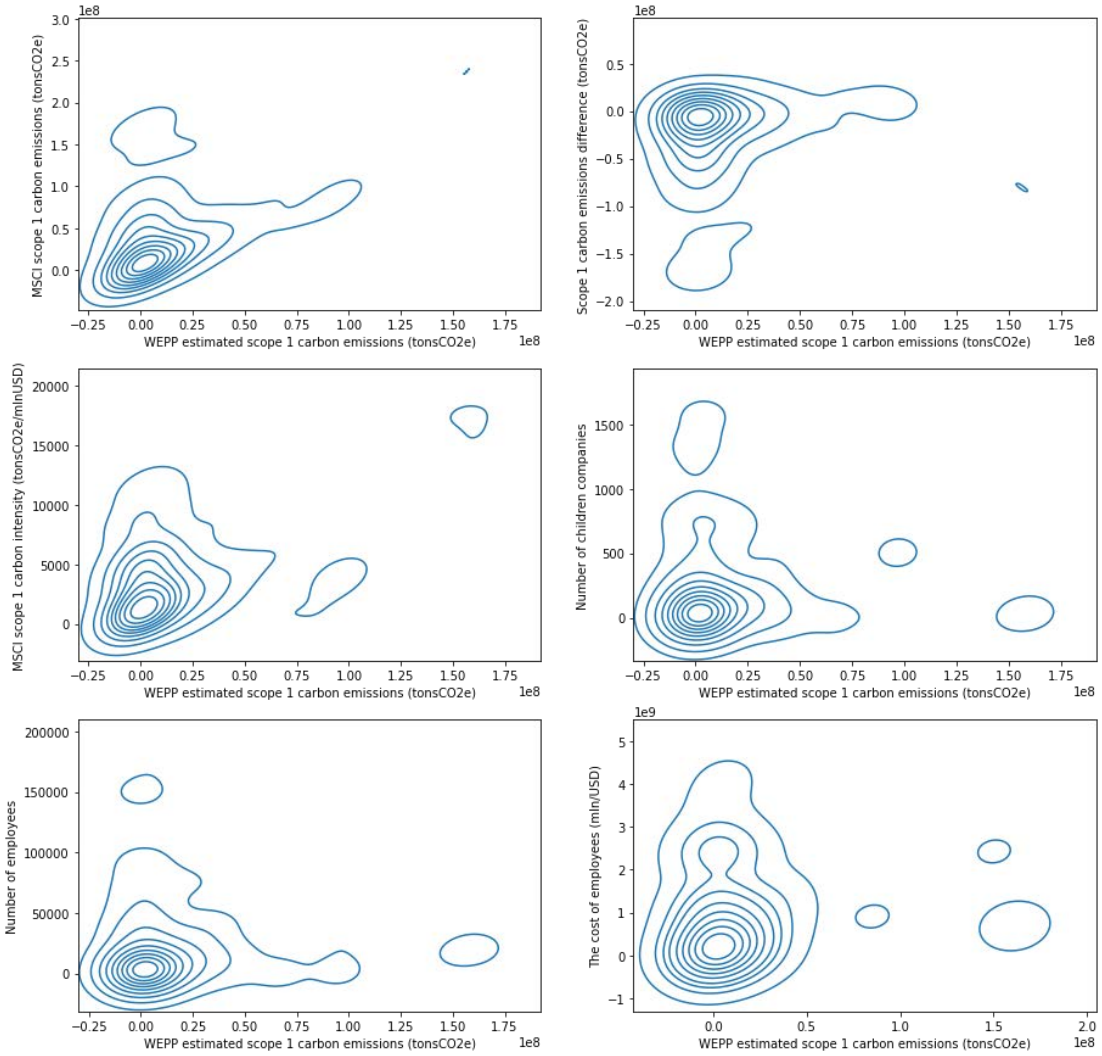

FIG. 36: The kernel density estimate (KDE) plots between WEPP estimated scope 1 carbon emissions (tonsCO_2e) and other quantitative factors, i.e., MSCI scope 1 carbon emissions (tonsCO_2e), scope 1 carbon emissions difference (tonsCO_2e), MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), number of children companies, number of employees, and the cost of employees (mln/USD).

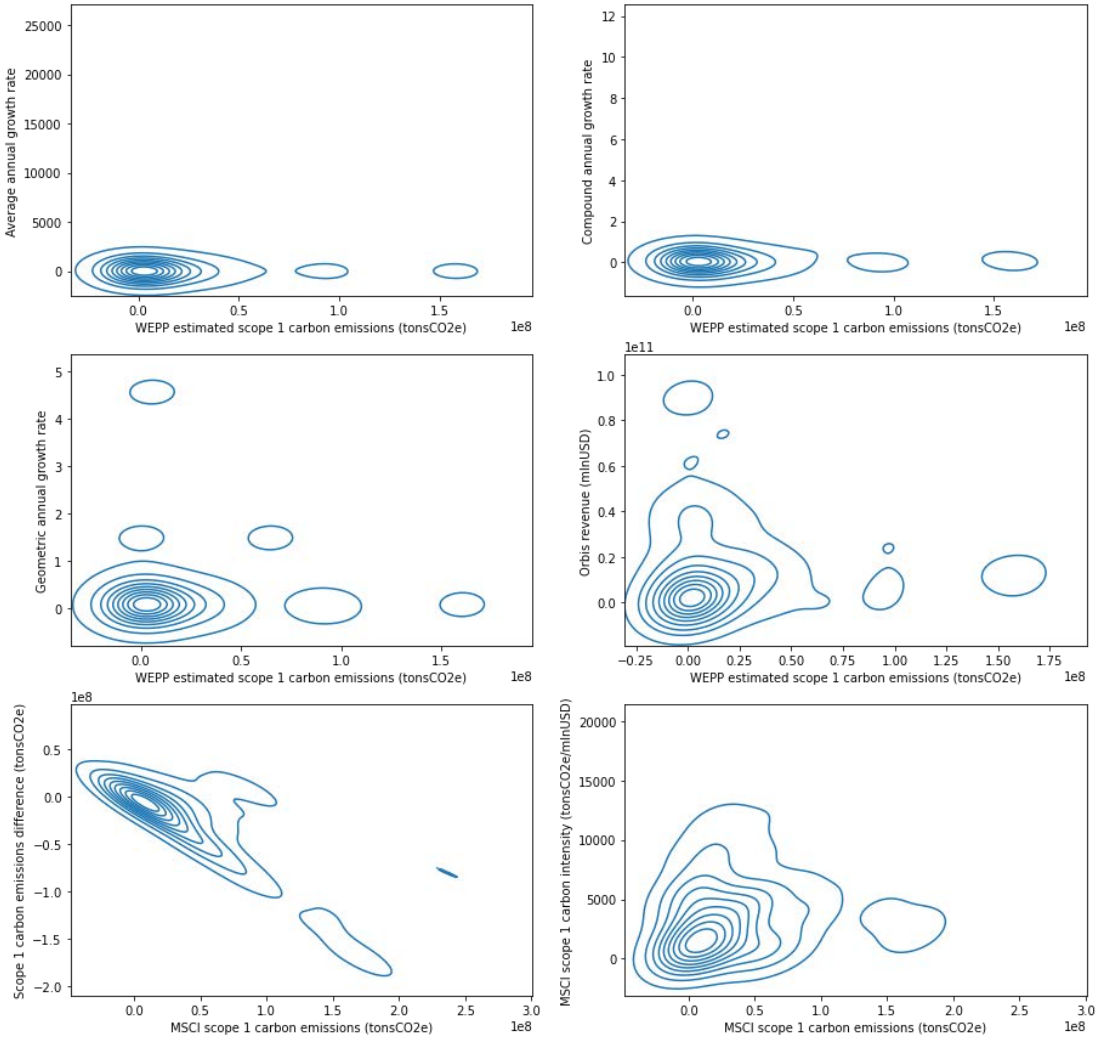

FIG. 37: The kernel density estimate (KDE) plots between WEPP estimated scope 1 carbon emissions (tonsCO_2e) and other quantitative factors, i.e., average annual growth rate, compound annual growth rate, geometric annual growth rate, and Orbis revenue (mln USD). And the kernel density estimate (KDE) plots between MSCI scope 1 carbon emissions (tonsCO_2e) and other quantitative factors, i.e., scope 1 carbon emissions difference (tonsCO_2e), and MSCI scope 1 carbon intensity (tonsCO_2e/mln USD).

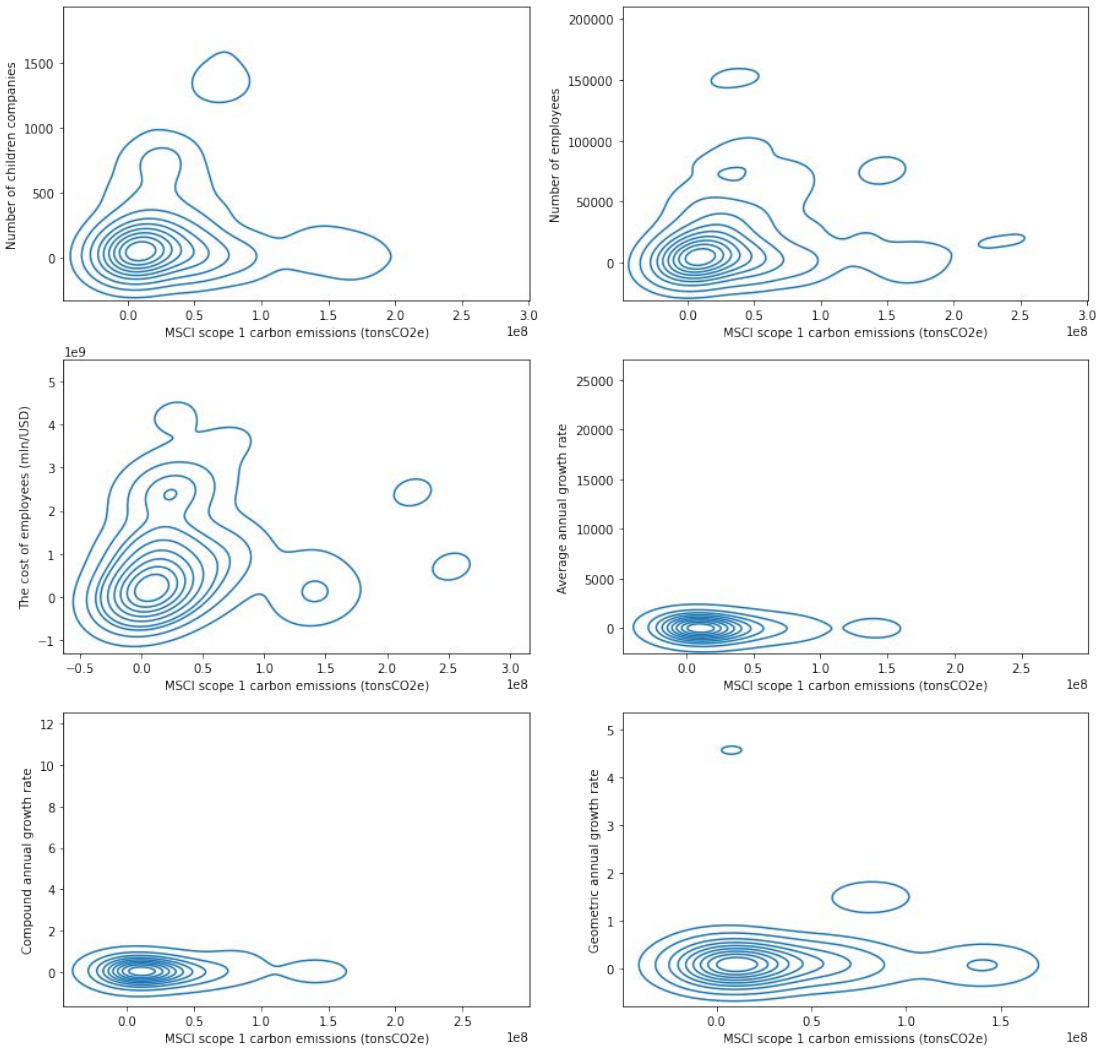

FIG. 38: The kernel density estimate (KDE) plots between MSCI scope 1 carbon emissions (tonsCO_2e) and other quantitative factors, i.e., number of children companies, number of employees, the cost of employees (mln/USD), average annual growth rate, compound annual growth rate, and geometric annual growth rate.

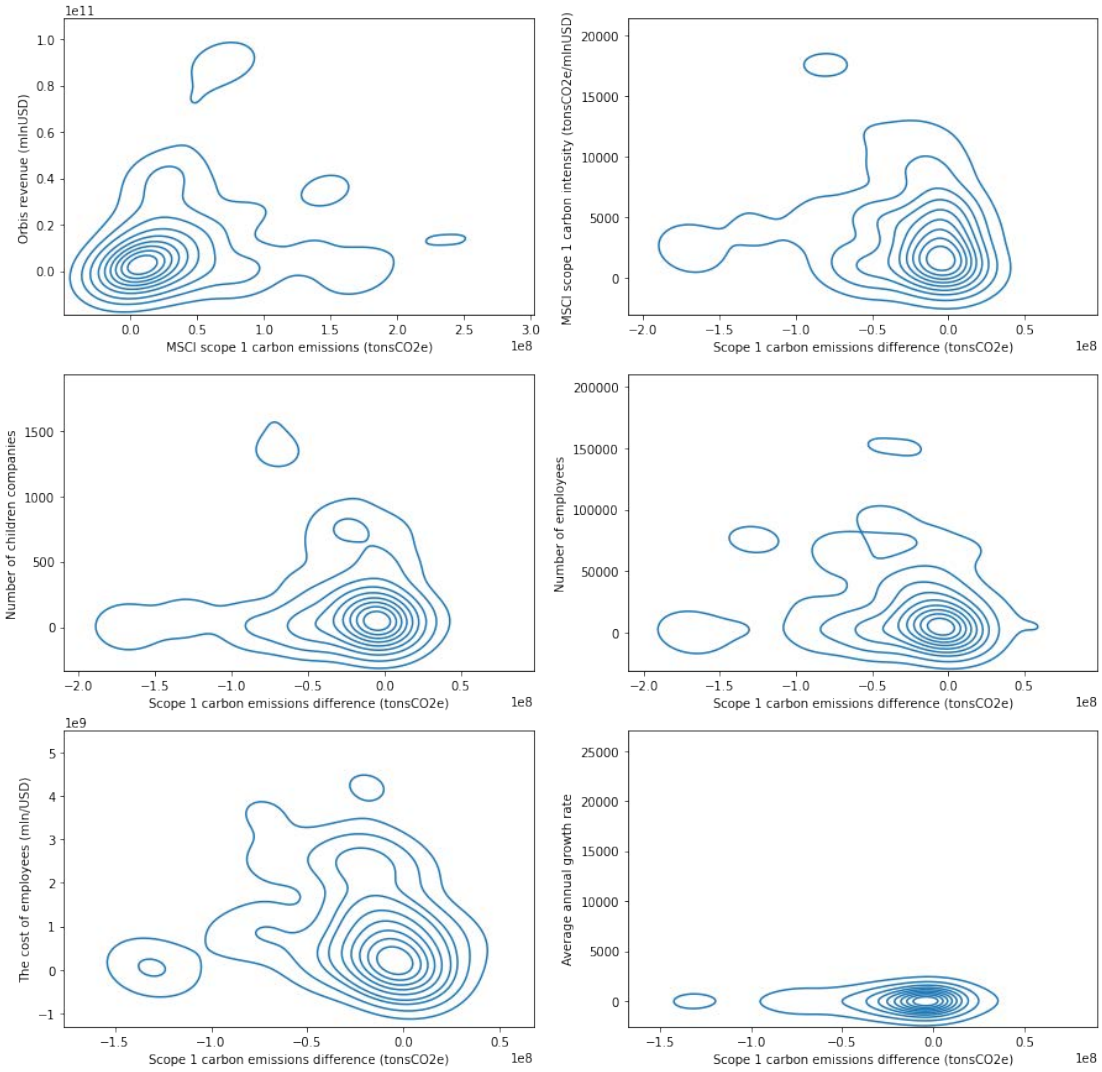

FIG. 39: The kernel density estimate (KDE) plot between MSCI scope 1 carbon emissions (tonsCO_2e) and Orbis revenue (mln USD). And the kernel density estimate (KDE) plots between scope 1 carbon emissions difference (tonsCO_2e) and other quantitative factors, i.e., MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), number of children companies, number of employees, the cost of employees (mln/USD), and average annual growth rate.

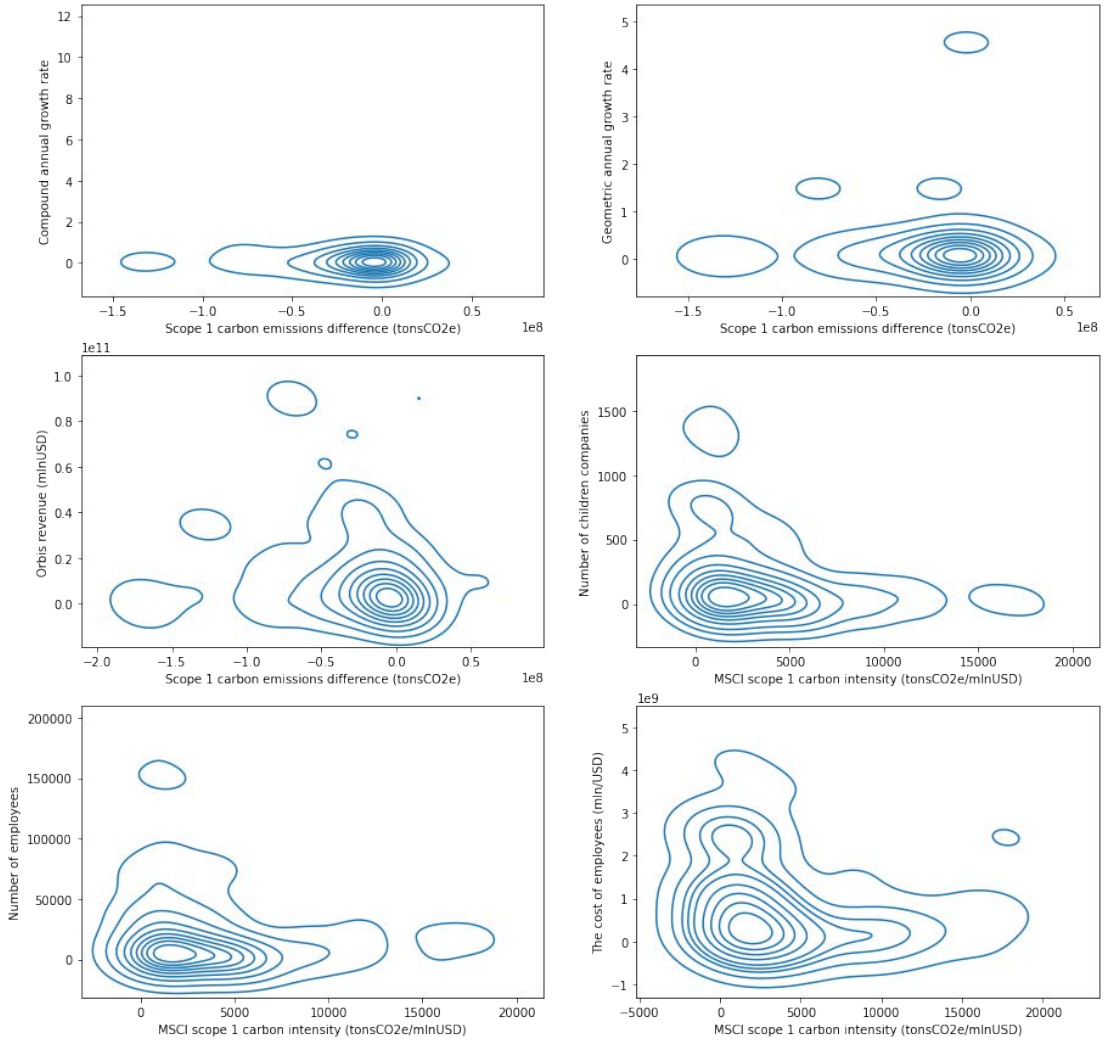

FIG. 40: The kernel density estimate (KDE) plots between scope 1 carbon emissions difference (tonsCO_2e) and other quantitative factors, i.e., compound annual growth rate, geometric annual growth rate, and Orbis revenue (mln USD). And the kernel density estimate (KDE) plots between MSCI scope 1 carbon intensity (tonsCO_2e/mln USD) and other quantitative factors, i.e., number of children companies, number of employees, and the cost of employees (mln/USD).

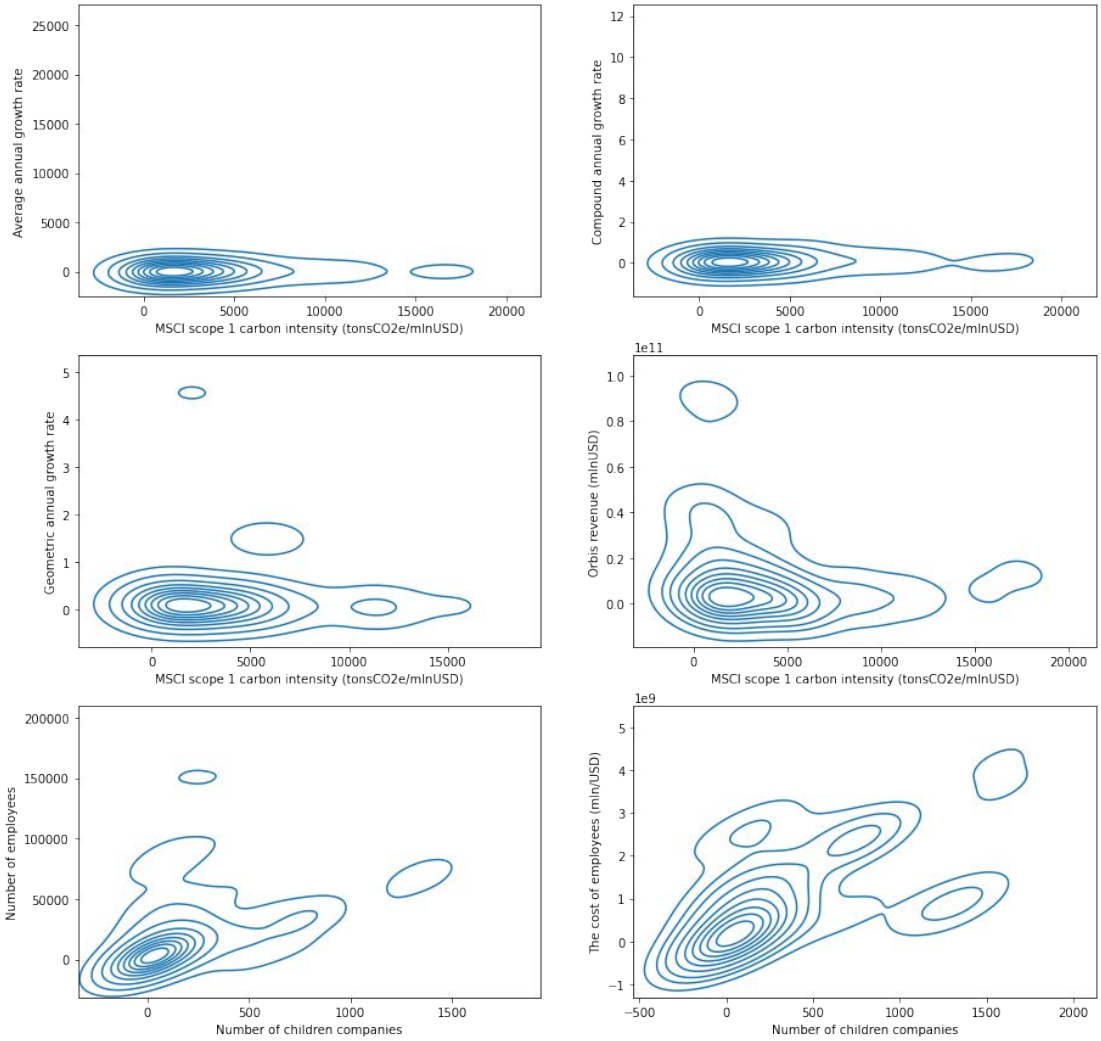

FIG. 41: The kernel density estimate (KDE) plots between MSCI scope 1 carbon intensity (tonsCO_2e/mln USD) and other quantitative factors, i.e., average annual growth rate, compound annual growth rate, geometric annual growth rate, and Orbis revenue (mln USD). And the kernel density estimate (KDE) plots between number of children companies and other quantitative factors, i.e., number of employees, and the cost of employees (mln/USD).

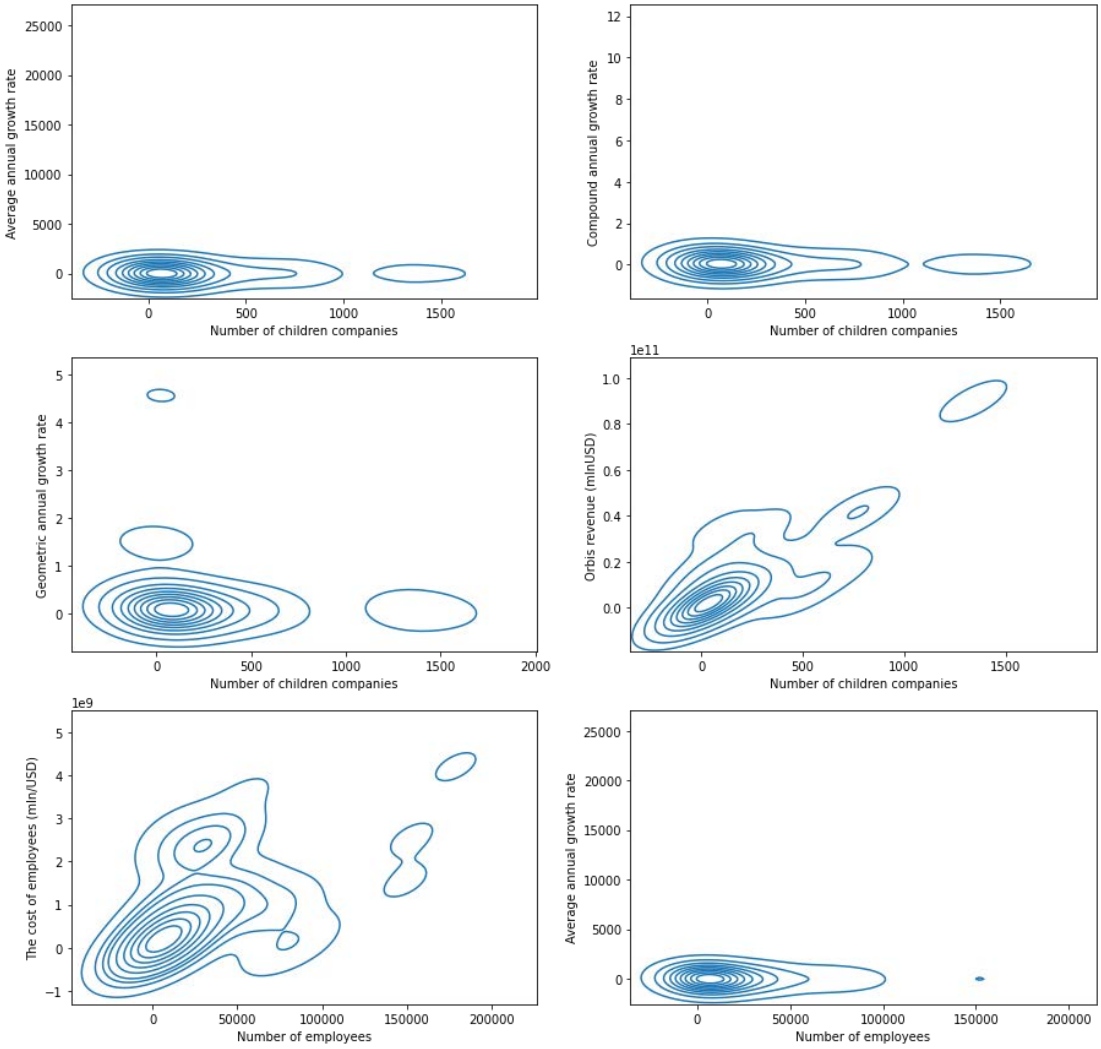

FIG. 42: The kernel density estimate (KDE) plots between number of children companies and other quantitative factors, i.e., average annual growth rate, compound annual growth rate, geometric annual growth rate, and Orbis revenue (mln USD). And the kernel density estimate (KDE) plots between number of employees and other quantitative factors, i.e., the cost of employees (mln/USD), and average annual growth rate.

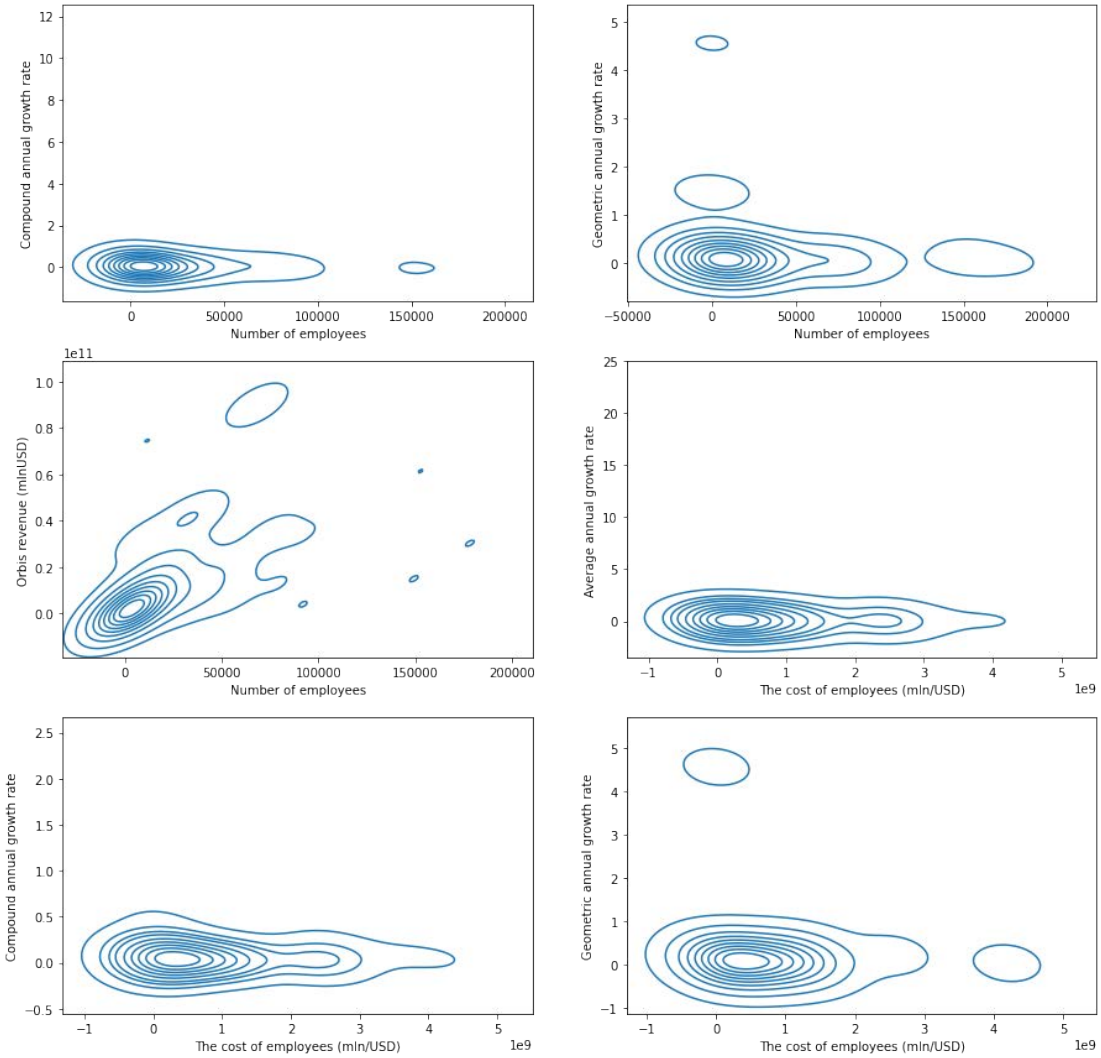

FIG. 43: The kernel density estimate (KDE) plots between number of employees and other quantitative factors, i.e., compound annual growth rate, and geometric annual growth rate, and Orbis revenue (mln USD). And the kernel density estimate (KDE) plots between the cost of employees (mln/USD) and other quantitative factors, i.e., average annual growth rate, compound annual growth rate, and geometric annual growth rate.

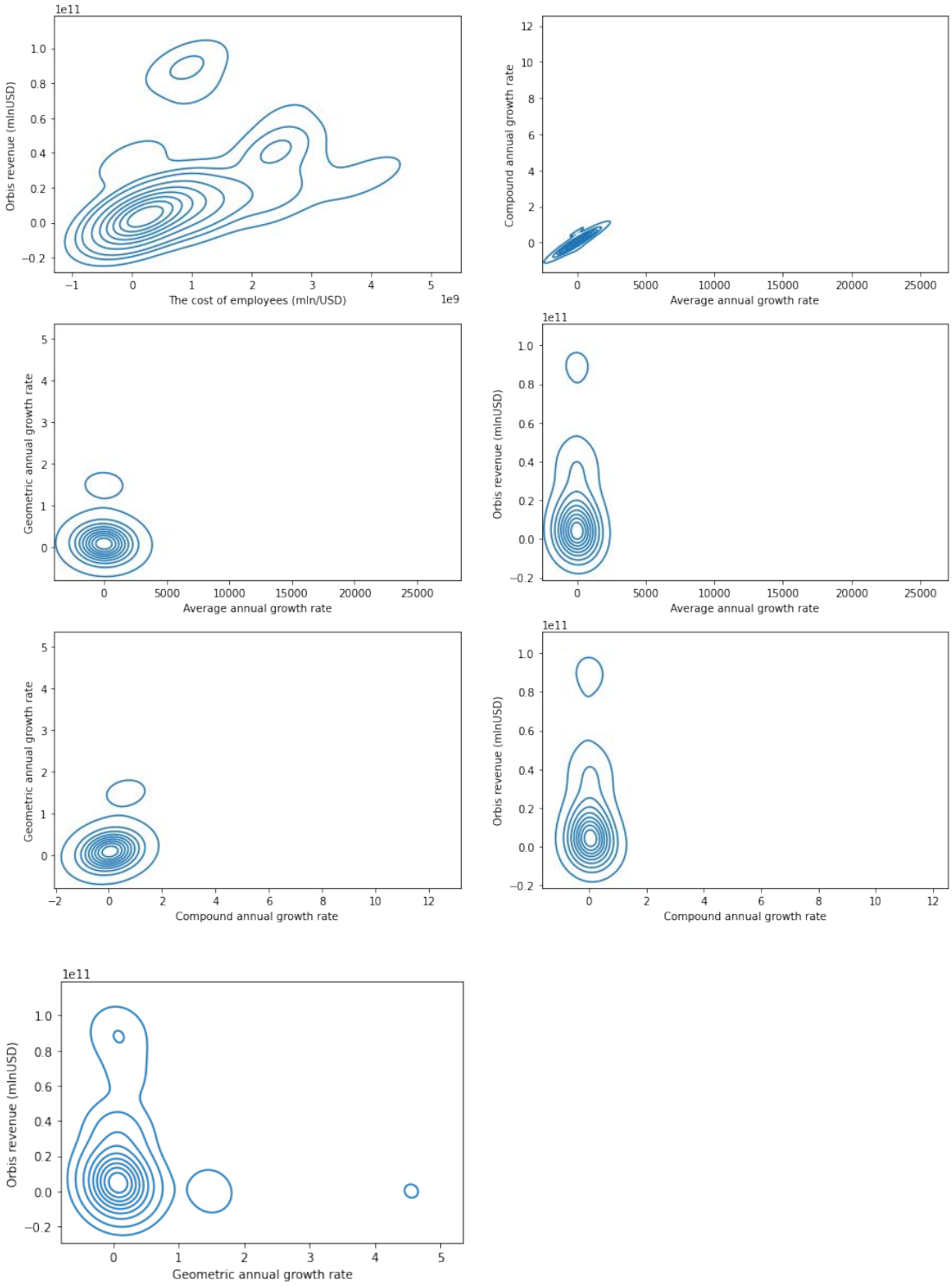

FIG. 44: The kernel density estimate (KDE) plot between number of employees and Orbis revenue (mln USD). The kernel density estimate (KDE) plots between average annual growth rate and other quantitative factors, i.e., compound annual growth rate, geometric annual growth rate, and Orbis revenue (mln USD). The kernel density estimate (KDE) plots between compound annual growth rate and other quantitative factors, i.e., geometric annual growth rate, and Orbis revenue (mln USD). The kernel density estimate (KDE) plot between geometric annual growth rate and Orbis revenue (mln USD).

APPENDIX B: THE HISTOGRAMS AND THE SCATTER PLOTS BETWEEN CDP DATASET AND MSCI (2021) DATASET

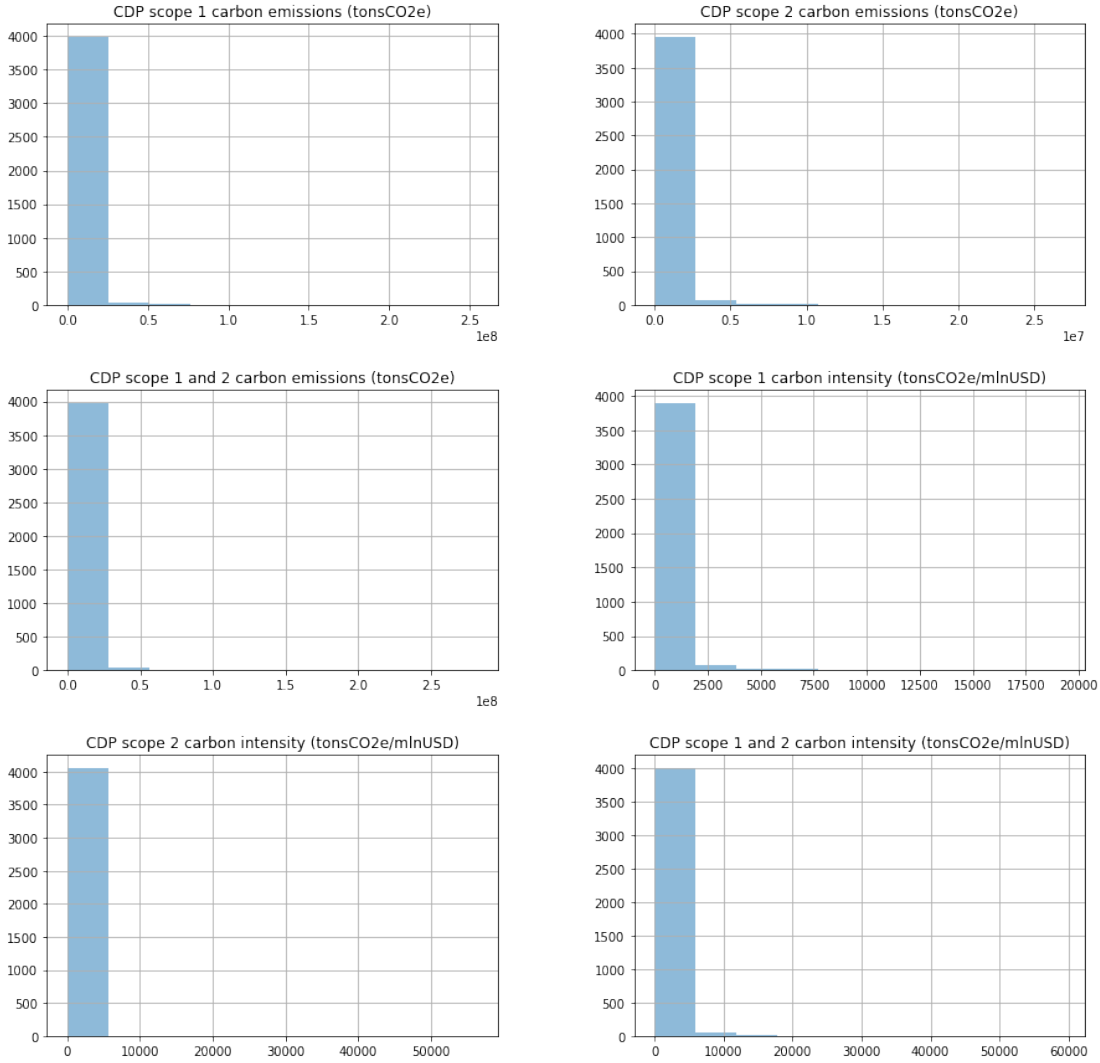

FIG. 45: The histograms of different types of carbon emissions and carbon intensity in CDP dataset, i.e., CDP scope 1 carbon emissions (tonsCO_2e), CDP scope 2 carbon emissions (tonsCO_2e), CDP scope 1 and 2 carbon emissions (tonsCO2e), CDP scope 1 carbon intensity (tonsCO_2e/mln USD), CDP scope 2 carbon intensity (tonsCO_2e/mln USD), and CDP scope 1 and 2 carbon intensity (tonsCO_2e/mlnUSD).

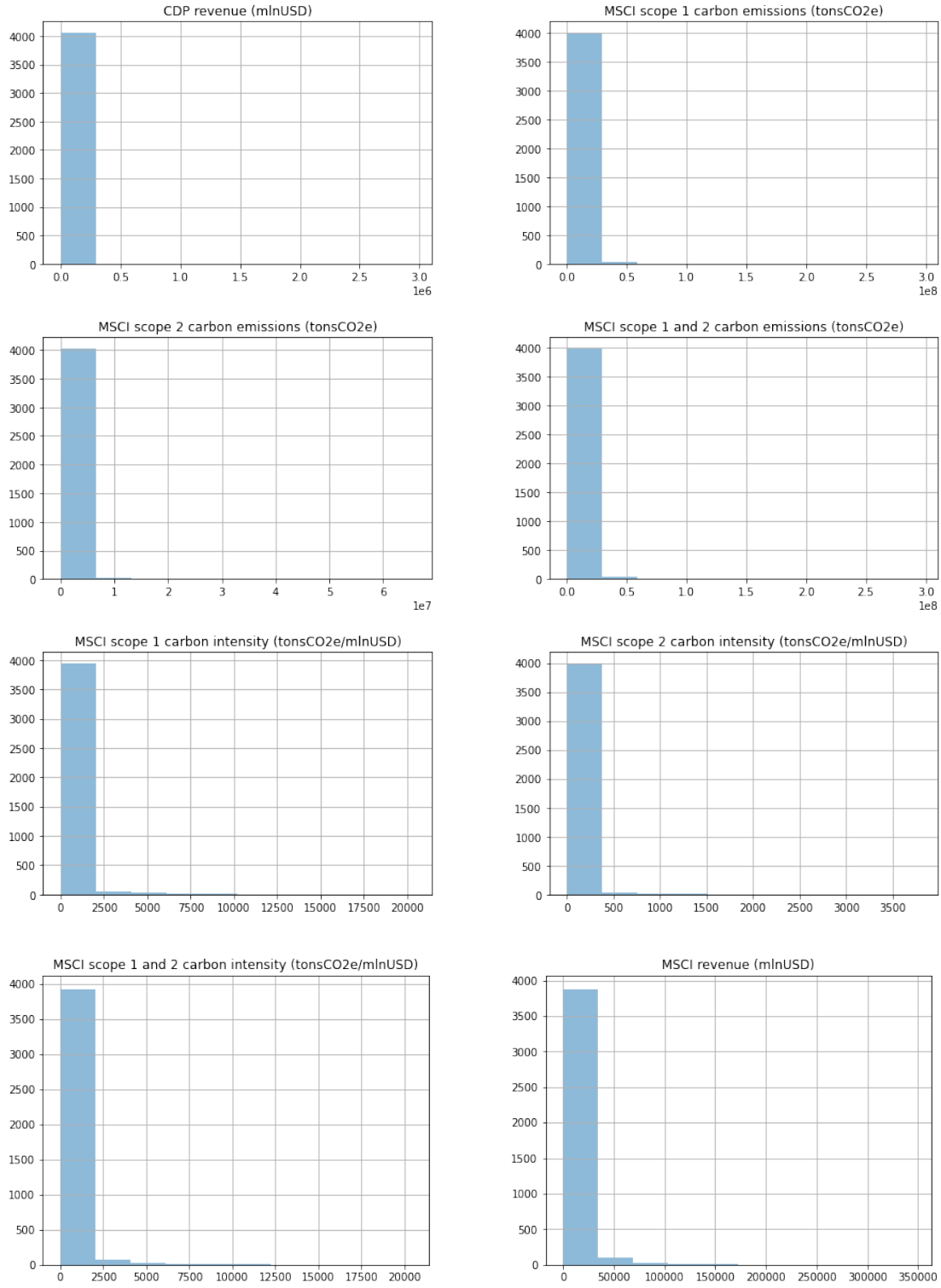

FIG. 46: The histograms of different types of carbon emissions and carbon intensity in CDP dataset and MSCI (2021) dataset, and the revenue, i.e., CDP revenue (mln USD), MSCI scope 1 carbon emissions (tonsCO_2e), MSCI scope 2 carbon emissions (tonsCO_2e), MSCI scope 1 and 2 carbon emissions (tonsCO_2e), MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), MSCI scope 2 carbon intensity (tonsCO_2e/mln USD), MSCI scope 1 and 2 carbon intensity (tonsCO_2e/mln USD), and MSCI revenue (mln USD).

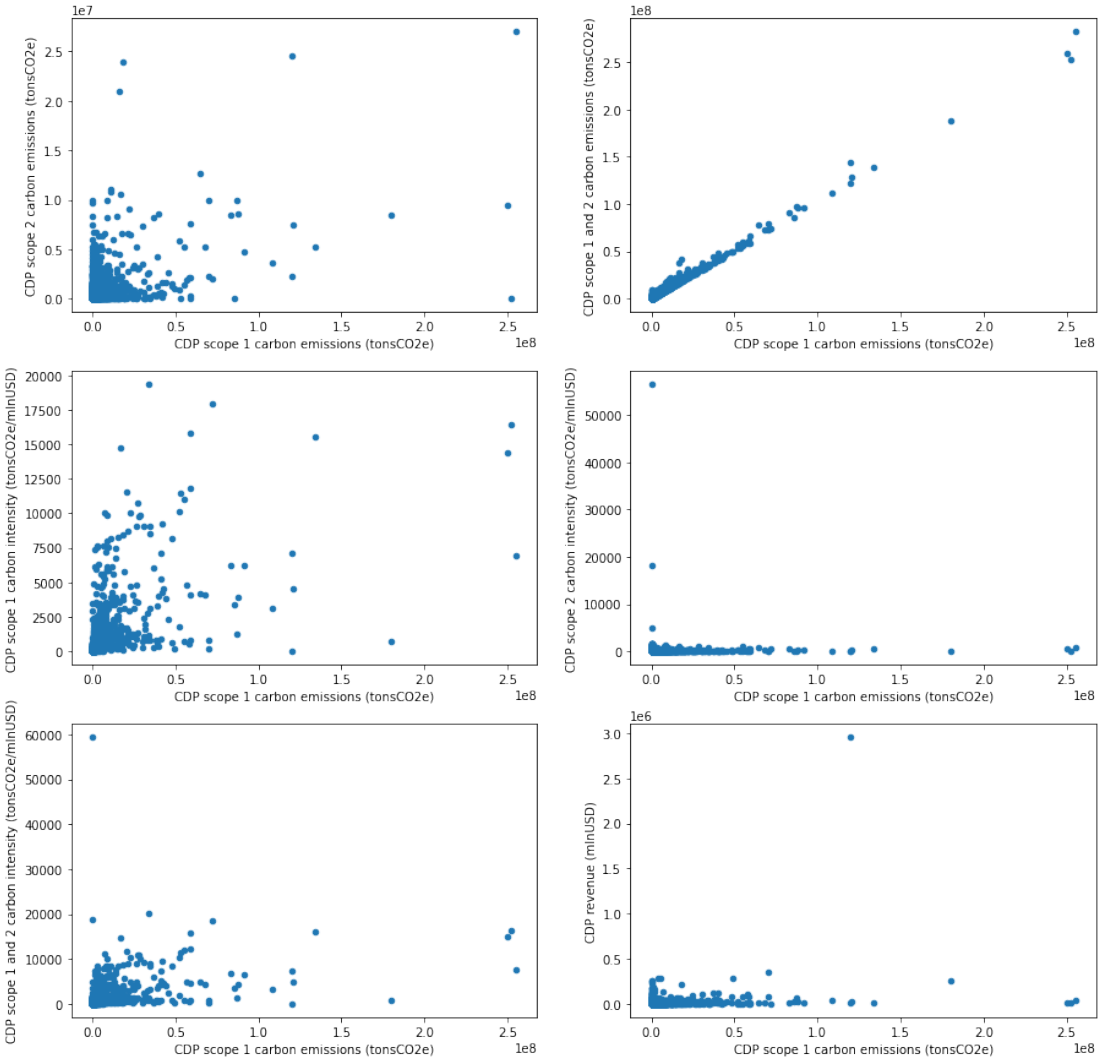

FIG. 47: The scatter plots between CDP scope 1 carbon emissions (tonsCO_2e) and other types of carbon emissions and carbon intensity and the revenue, i.e., CDP scope 2 carbon emissions (tonsCO_2e), CDP scope 1 and 2 carbon emissions (tonsCO_2e), CDP scope 1 carbon intensity (tonsCO_2e/mlnUSD), CDP scope 2 carbon intensity (tonsCO_2e/mln USD), CDP scope 1 and 2 carbon intensity (tonsCO_2e/mln USD), and CDP revenue (mln USD).

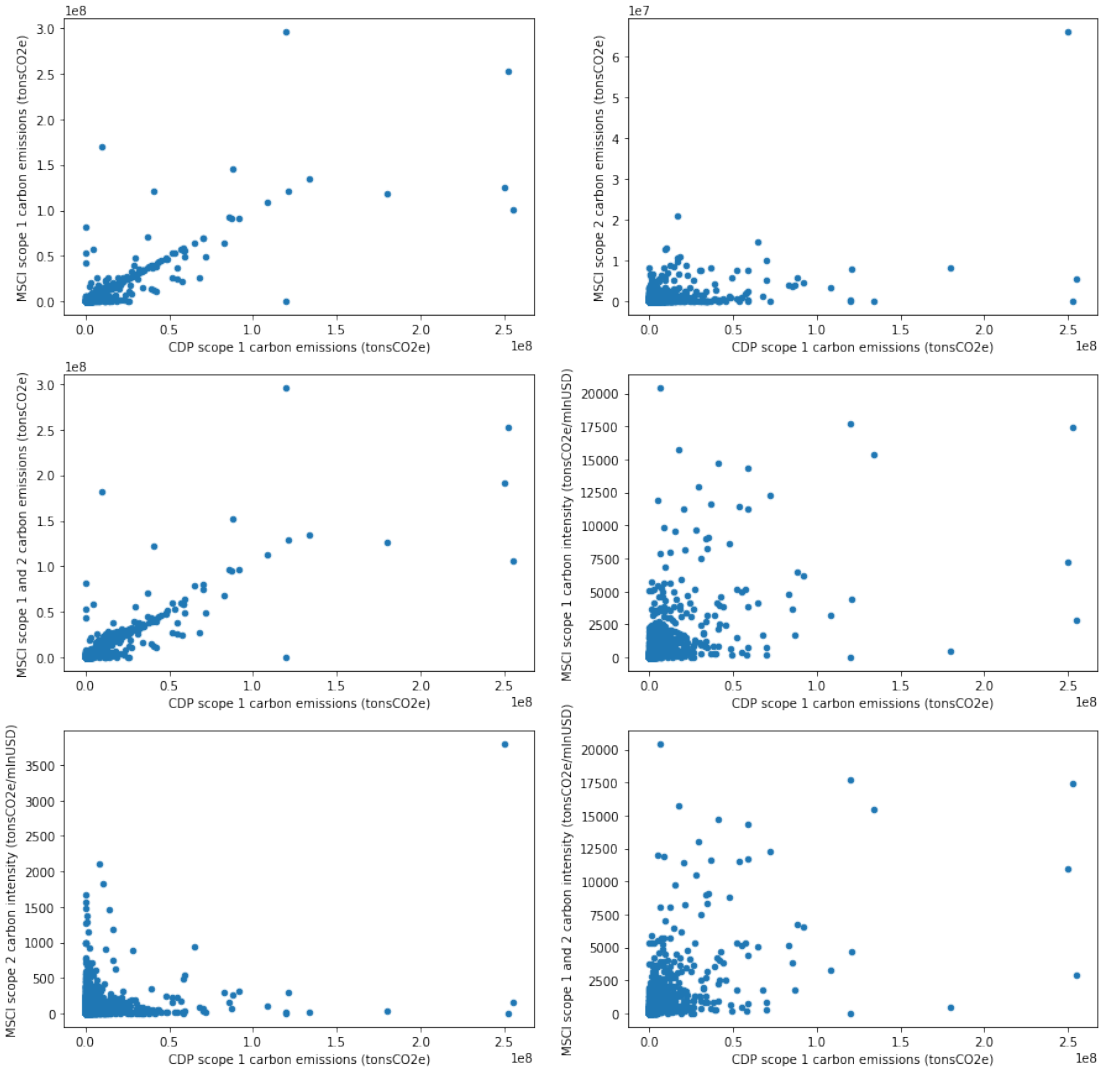

FIG. 48: The scatter plots between CDP scope 1 carbon emissions (tonsCO_2e) and other types of carbon emissions and carbon intensity, i.e., MSCI scope 1 carbon emissions (tonsCO_2e), MSCI scope 2 carbon emissions (tonsCO_2e), MSCI scope 1 and 2 carbon emissions (tonsCO_2e), MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), MSCI scope 2 carbon intensity (tonsCO_2e/mln USD), and MSCI scope 1 and 2 carbon intensity (tonsCO_2e/mln USD).

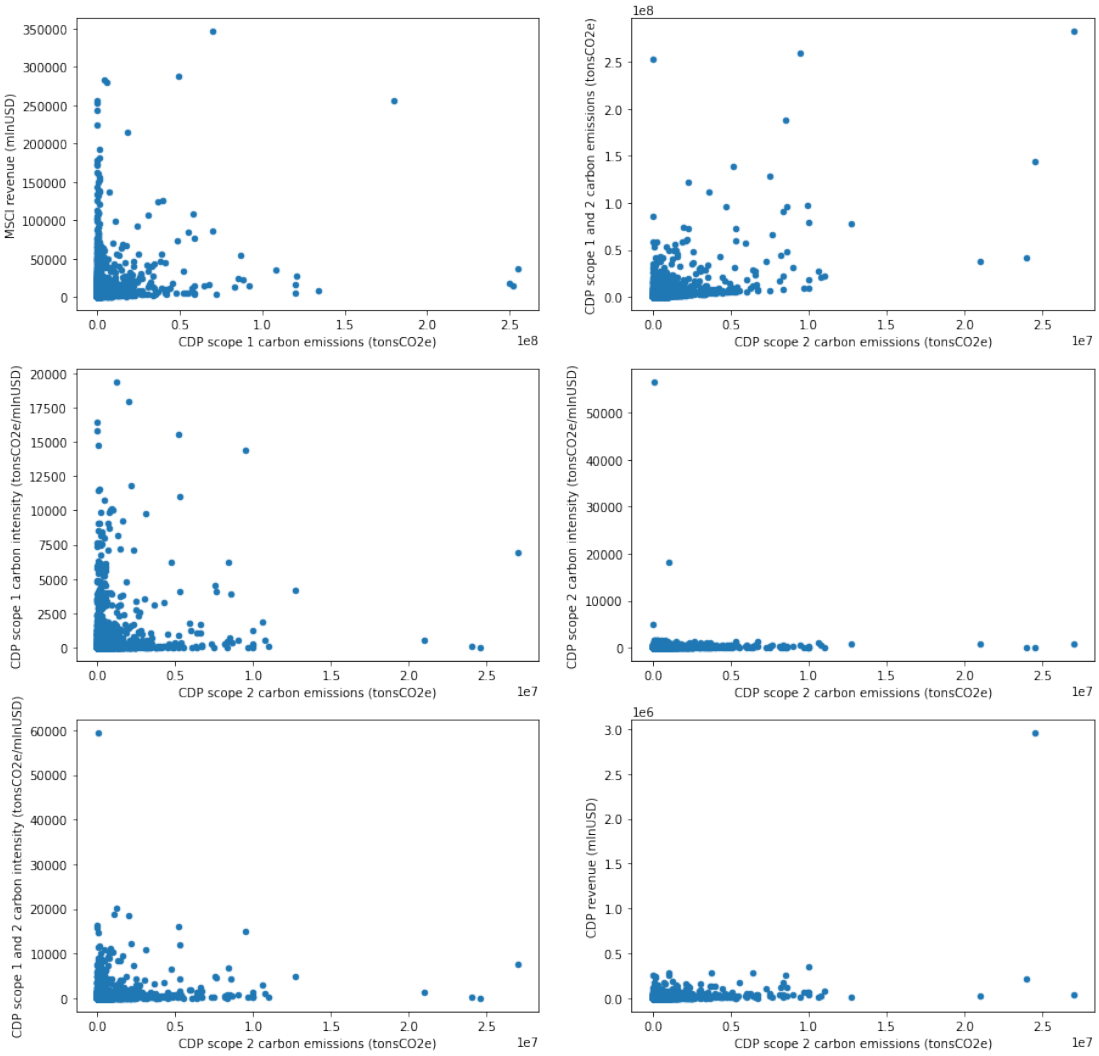

FIG. 49: The scatter plot between CDP scope 1 carbon emissions (tonsCO_2e) and MSCI revenue (mln USD). The scatter plots between CDP scope 2 carbon emissions (tonsCO_2e) and other types of carbon emissions and carbon intensity and the revenue, i.e., CDP scope 1 and 2 carbon emissions (tonsCO_2e), CDP scope 1 carbon intensity (tonsCO_2e/mlnUSD), CDP scope 2 carbon intensity (tonsCO_2e/mln USD), CDP scope 1 and 2 carbon intensity (tonsCO_2e/mln USD), and CDP revenue (mln USD).

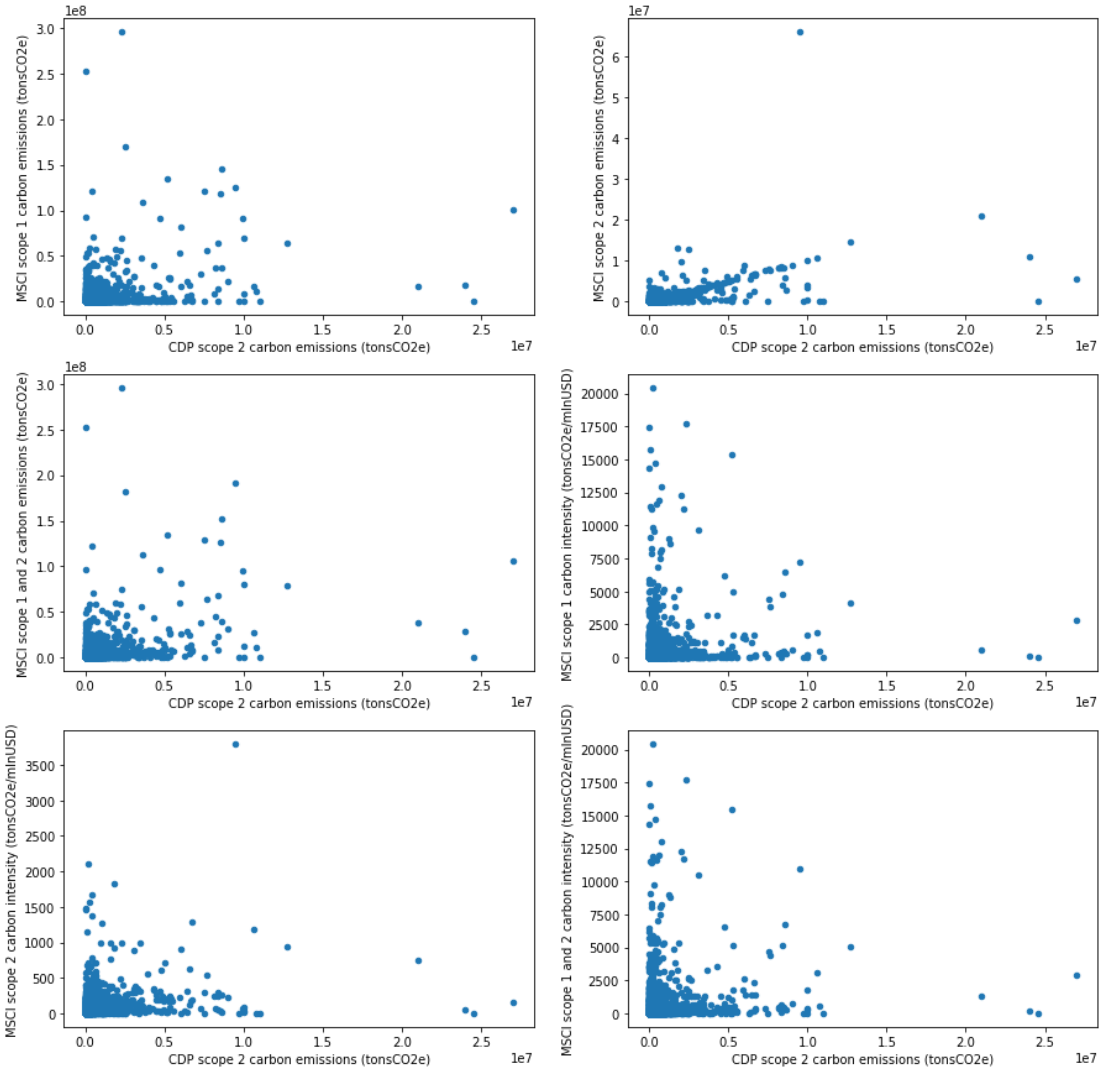

FIG. 50: The scatter plots between CDP scope 2 carbon emissions (tonsCO_2e) and other types of carbon emissions and carbon intensity, i.e., MSCI scope 1 carbon emissions (tonsCO_2e), MSCI scope 2 carbon emissions (tonsCO_2e), MSCI scope 1 and 2 carbon emissions (tonsCO_2e), MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), MSCI scope 2 carbon intensity (tonsCO_2e/mln USD), and MSCI scope 1 and 2 carbon intensity (tonsCO_2e/mln USD).

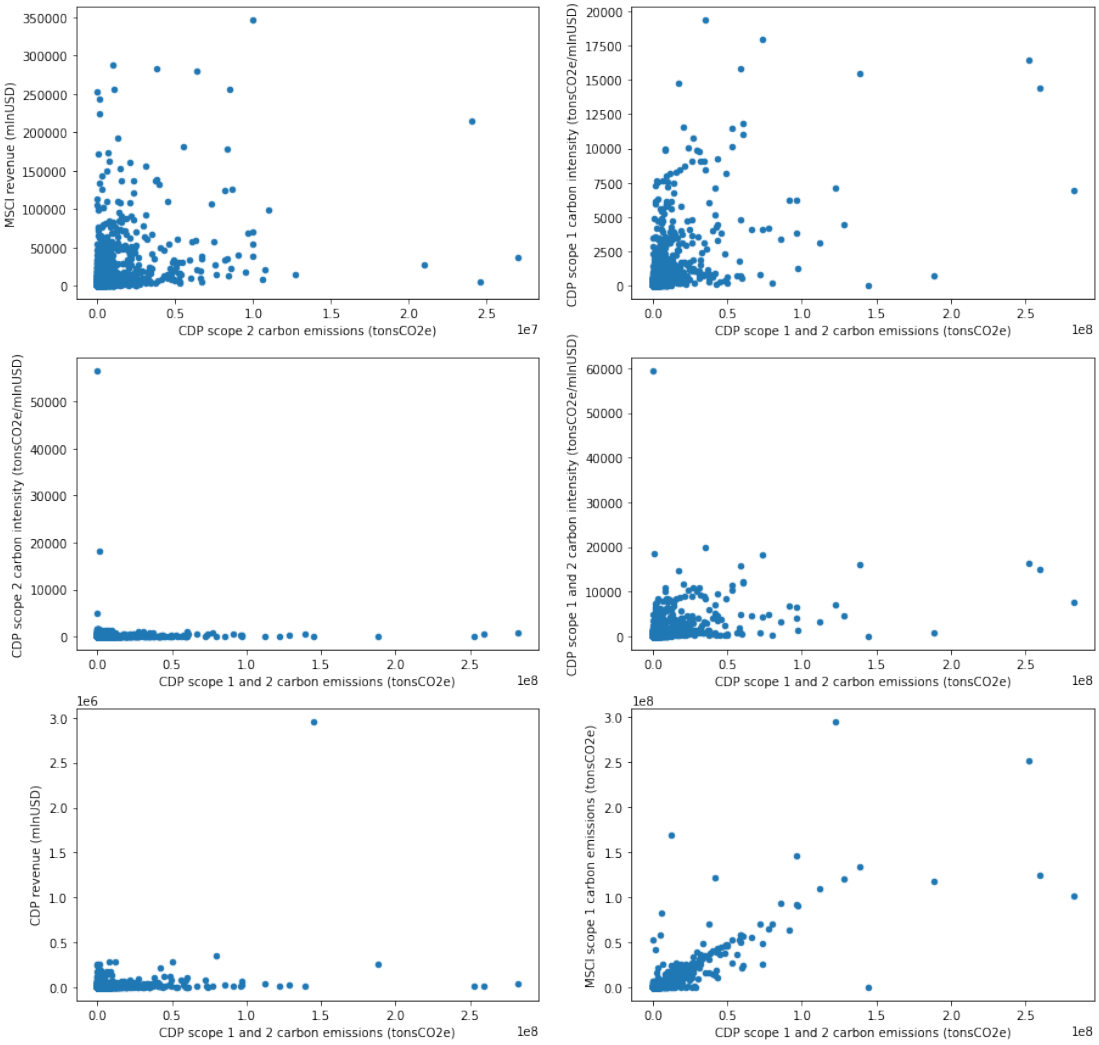

FIG. 51: The scatter plot between CDP scope 2 carbon emissions (tonsCO_2e) and MSCI revenue (mln USD). The scatter plots between CDP scope 1 and 2 carbon emissions (tonsCO_2e) and other types of carbon emissions and carbon intensity, i.e., CDP scope 1 carbon intensity (tonsCO_2e/mlnUSD), CDP scope 2 carbon intensity (tonsCO_2e/mln USD), CDP scope 1 and 2 carbon intensity (tonsCO_2e/mln USD), CDP revenue (mln USD), and MSCI scope 1 carbon emissions (tonsCO_2e).

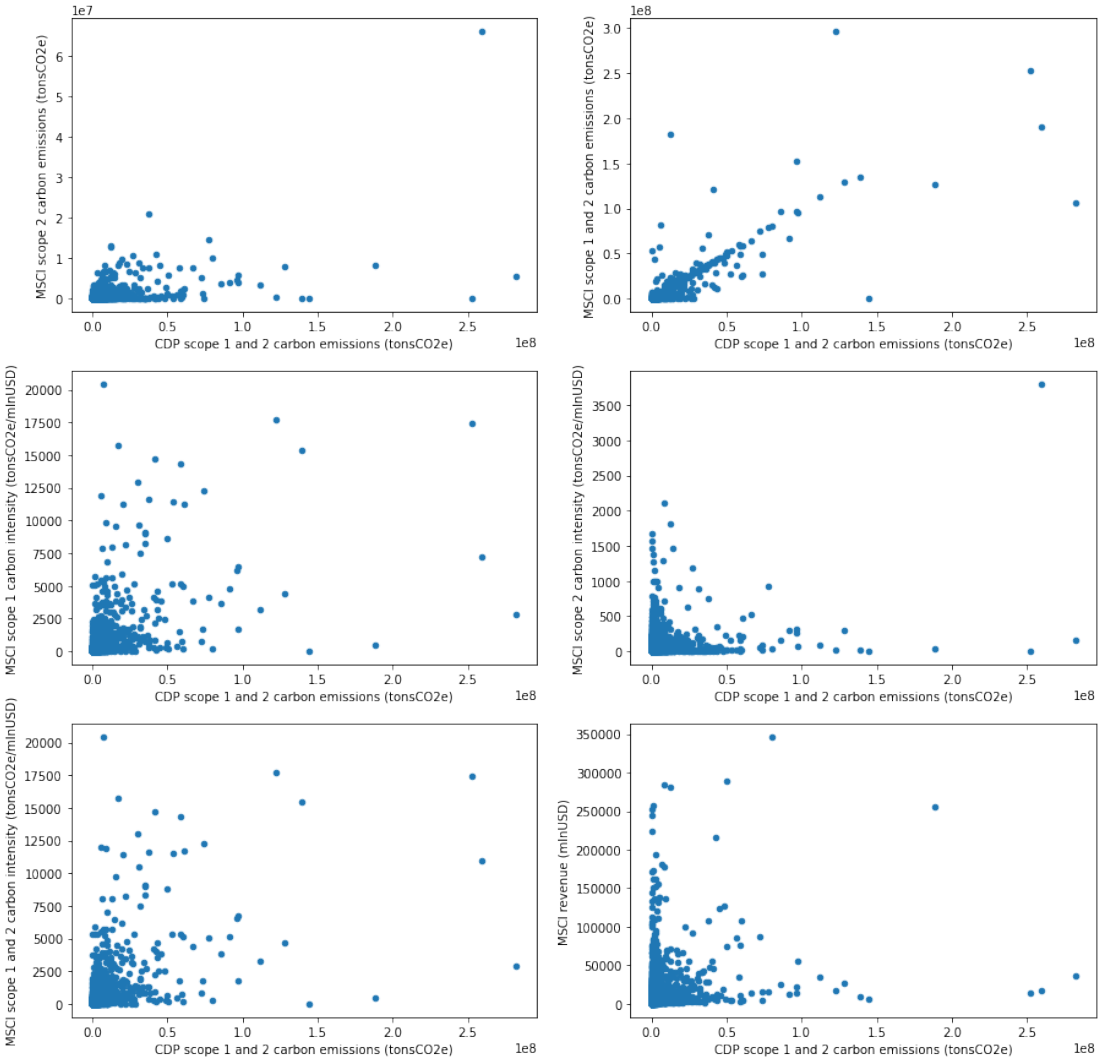

FIG. 52: The scatter plots between CDP scope 1 and 2 carbon emissions (tonsCO_2e) and other types of carbon emissions and carbon intensity and the revenue, i.e., MSCI scope 2 carbon emissions (tonsCO_2e), MSCI scope 1 and 2 carbon emissions (tonsCO_2e), MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), MSCI scope 2 carbon intensity (tonsCO_2e/mln USD), MSCI scope 1 and 2 carbon intensity (tonsCO_2e/mln USD), and MSCI revenue (mln USD).

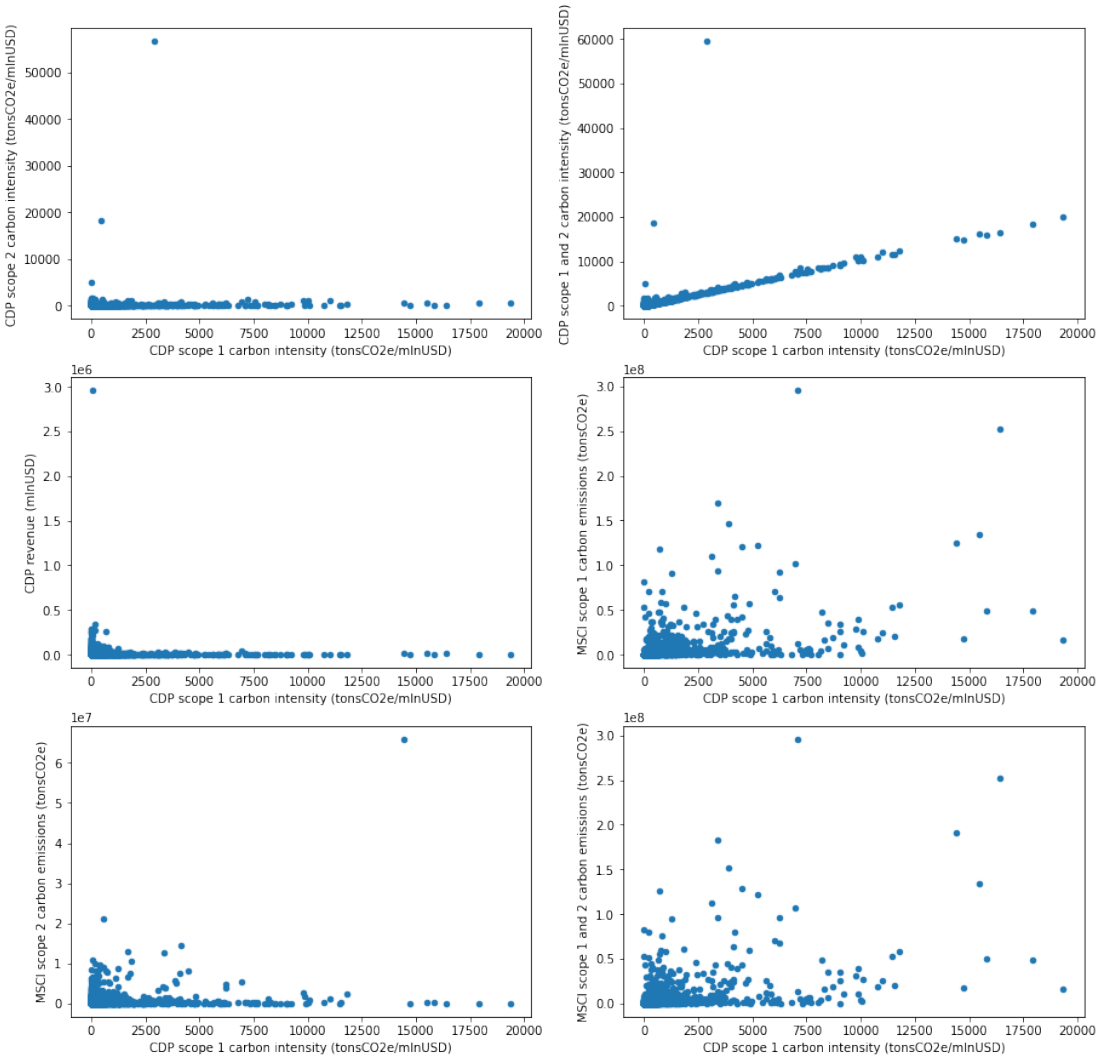

FIG. 53: The scatter plots between CDP scope 1 carbon intensity (tonsCO_2e/mlnUSD) and other types of carbon emissions and carbon intensity, i.e., CDP scope 2 carbon intensity (tonsCO_2e/mln USD), CDP scope 1 and 2 carbon intensity (tonsCO_2e/mln USD), CDP revenue (mln USD), MSCI scope 1 carbon emissions (tonsCO_2e), MSCI scope 2 carbon emissions (tonsCO_2e), and MSCI scope 1 and 2 carbon emissions (tonsCO_2e).

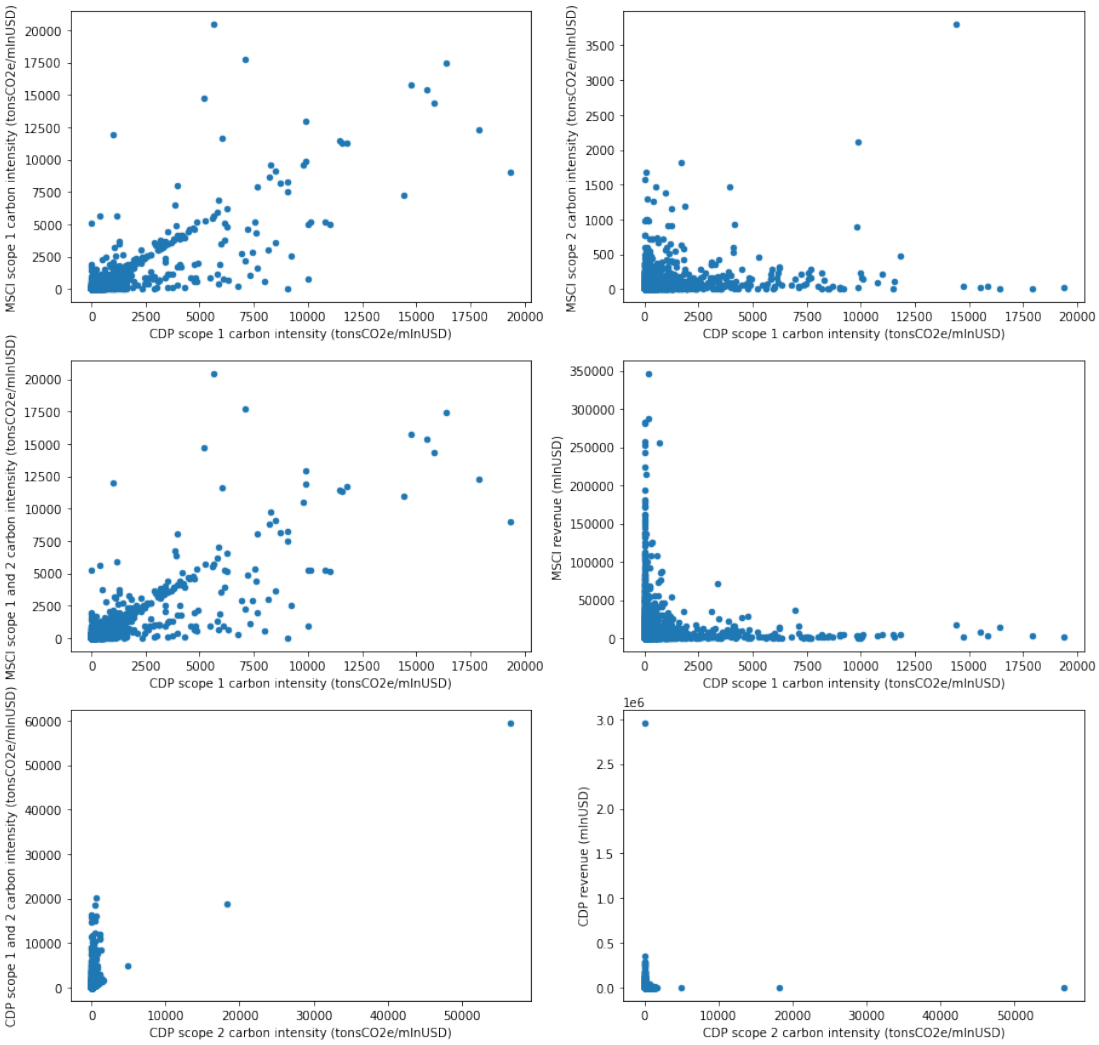

FIG. 54: The scatter plots between CDP scope 1 carbon intensity (tonsCO_2e/mlnUSD) and other types of carbon emissions and carbon intensity and the revenue, i.e., MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), MSCI scope 2 carbon intensity (tonsCO_2e/mln USD), MSCI scope 1 and 2 carbon intensity (tonsCO_2e/mln USD), and MSCI revenue (mln USD). And the scatter plots between CDP scope 2 carbon intensity (tonsCO_2e/mln USD) and other types of carbon emissions and carbon intensity and the revenue, i.e., CDP scope 1 and 2 carbon intensity (tonsCO_2e/mln USD), and CDP revenue (mln USD).

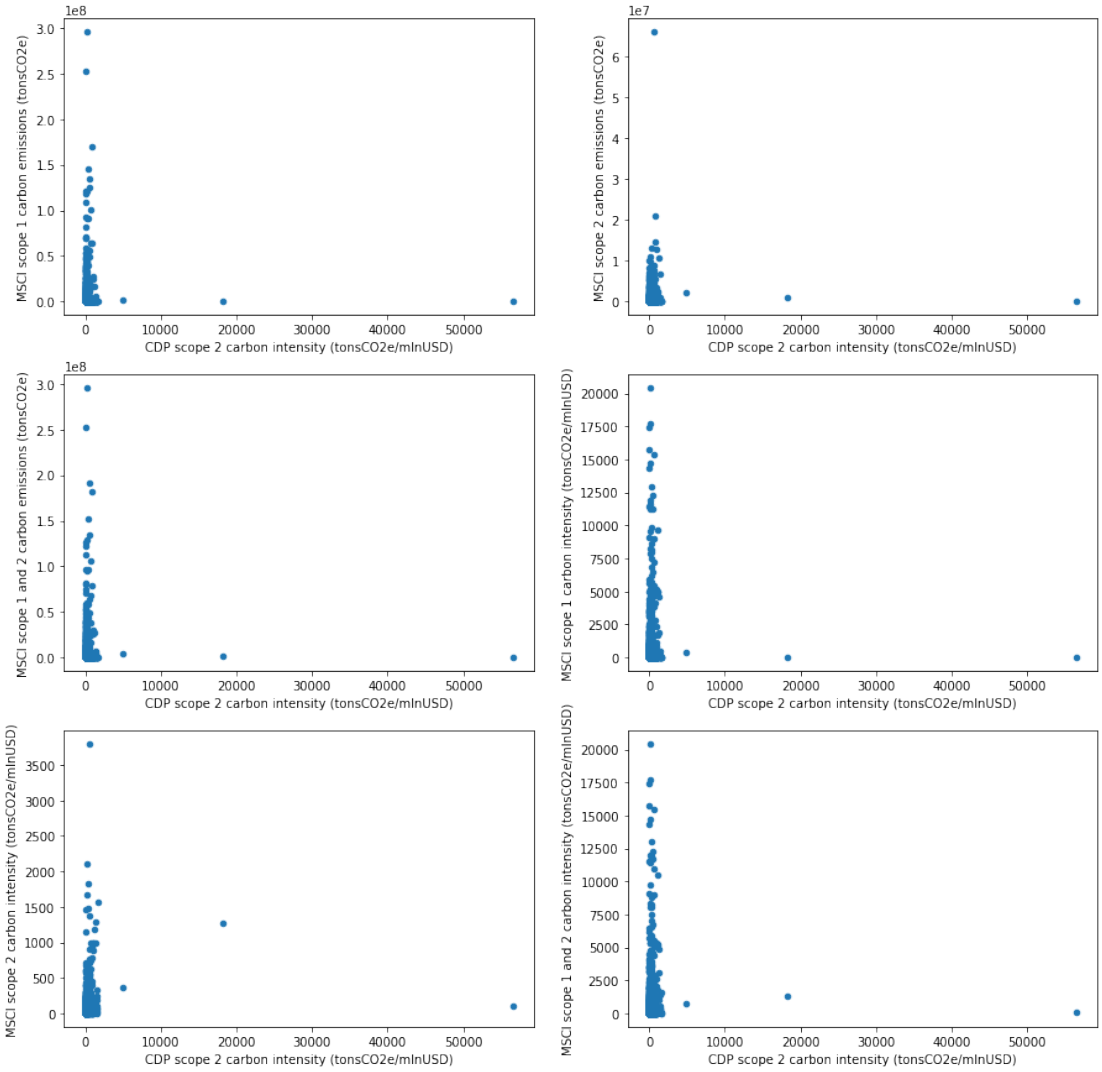

FIG. 55: The scatter plots between CDP scope 2 carbon intensity (tonsCO_2e/mln USD) and other types of carbon emissions and carbon intensity, i.e., MSCI scope 1 carbon emissions (tonsCO_2e), MSCI scope 2 carbon emissions (tonsCO_2e), MSCI scope 1 and 2 carbon emissions (tonsCO_2e), MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), MSCI scope 2 carbon intensity (tonsCO_2e/mln USD), and MSCI scope 1 and 2 carbon intensity (tonsCO_2e/mln USD).

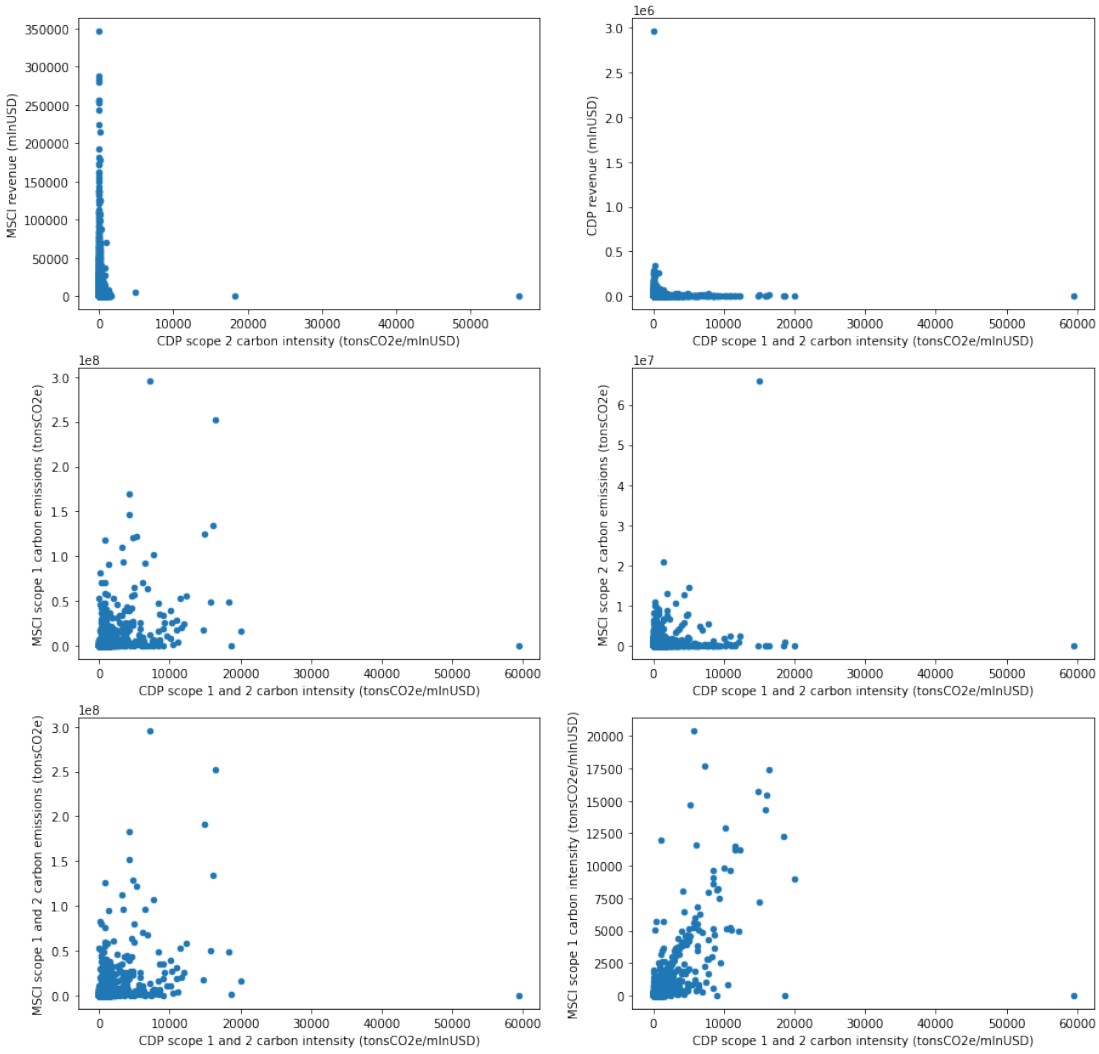

FIG. 56: The scatter plot between CDP scope 2 carbon intensity (tonsCO_2e/mln USD) and MSCI revenue (mln USD). The scatter plots between CDP scope 1 and 2 carbon intensity (tonsCO_2e/mln USD) and other types of carbon emissions and carbon intensity and revenue, i.e., CDP revenue (mln USD), MSCI scope 1 carbon emissions (tonsCO_2e), MSCI scope 2 carbon emissions (tonsCO_2e), MSCI scope 1 and 2 carbon emissions (tonsCO_2e), and MSCI scope 1 carbon intensity (tonsCO_2e/mln USD).

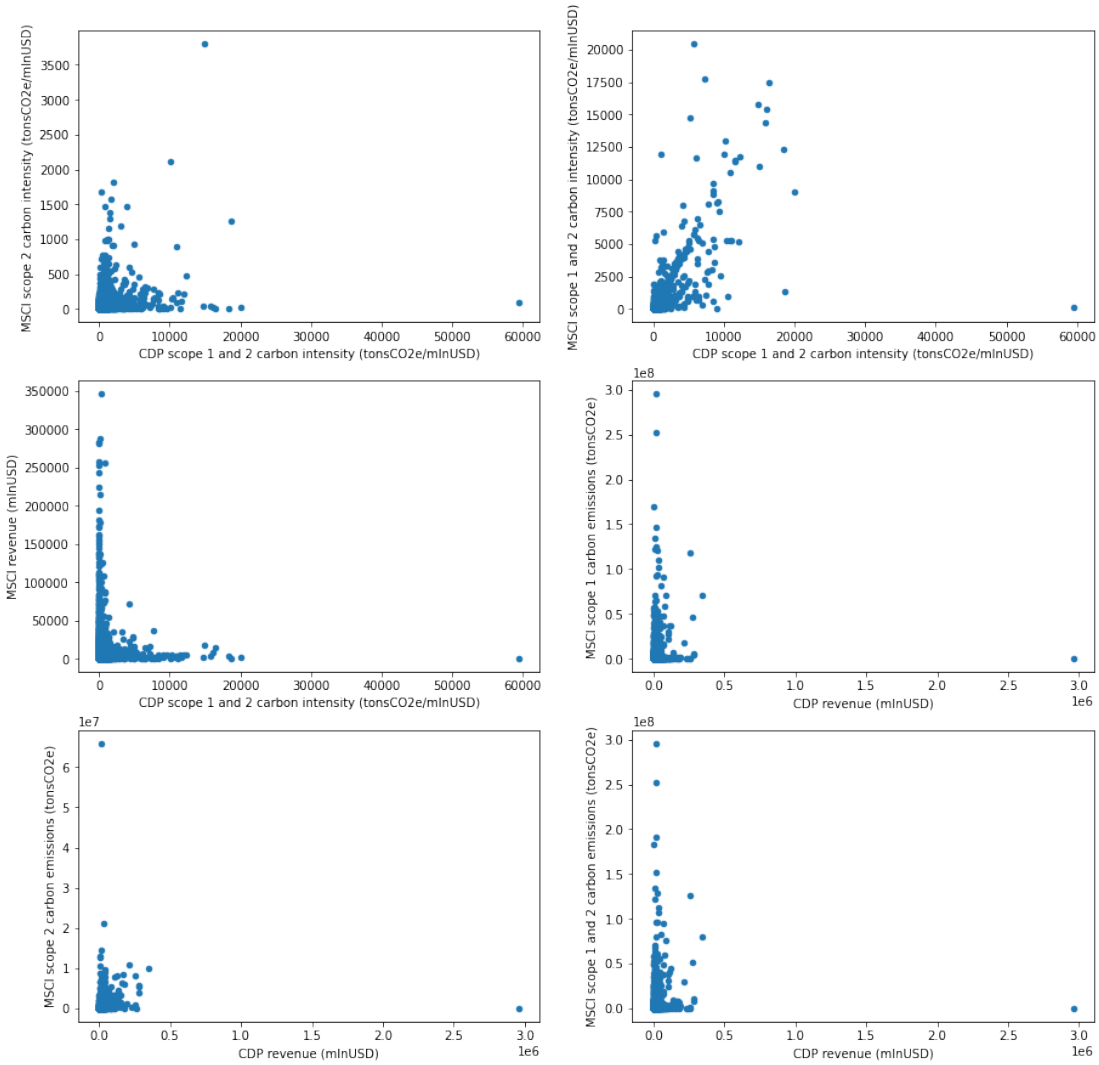

FIG. 57: The scatter plots between CDP scope 1 and 2 carbon intensity (tonsCO_2e/mln USD) and other types of carbon emissions and carbon intensity and revenue, i.e., MSCI scope 2 carbon intensity (tonsCO_2e/mln USD), MSCI scope 1 and 2 carbon intensity (tonsCO_2e/mln USD), and MSCI revenue (mln USD). The scatter plots between CDP revenue (mln USD) and other types of carbon emissions and carbon intensity, i.e., MSCI scope 1 carbon emissions (tonsCO_2e), MSCI scope 2 carbon emissions (tonsCO_2e), and MSCI scope 1 and 2 carbon emissions (tonsCO_2e).

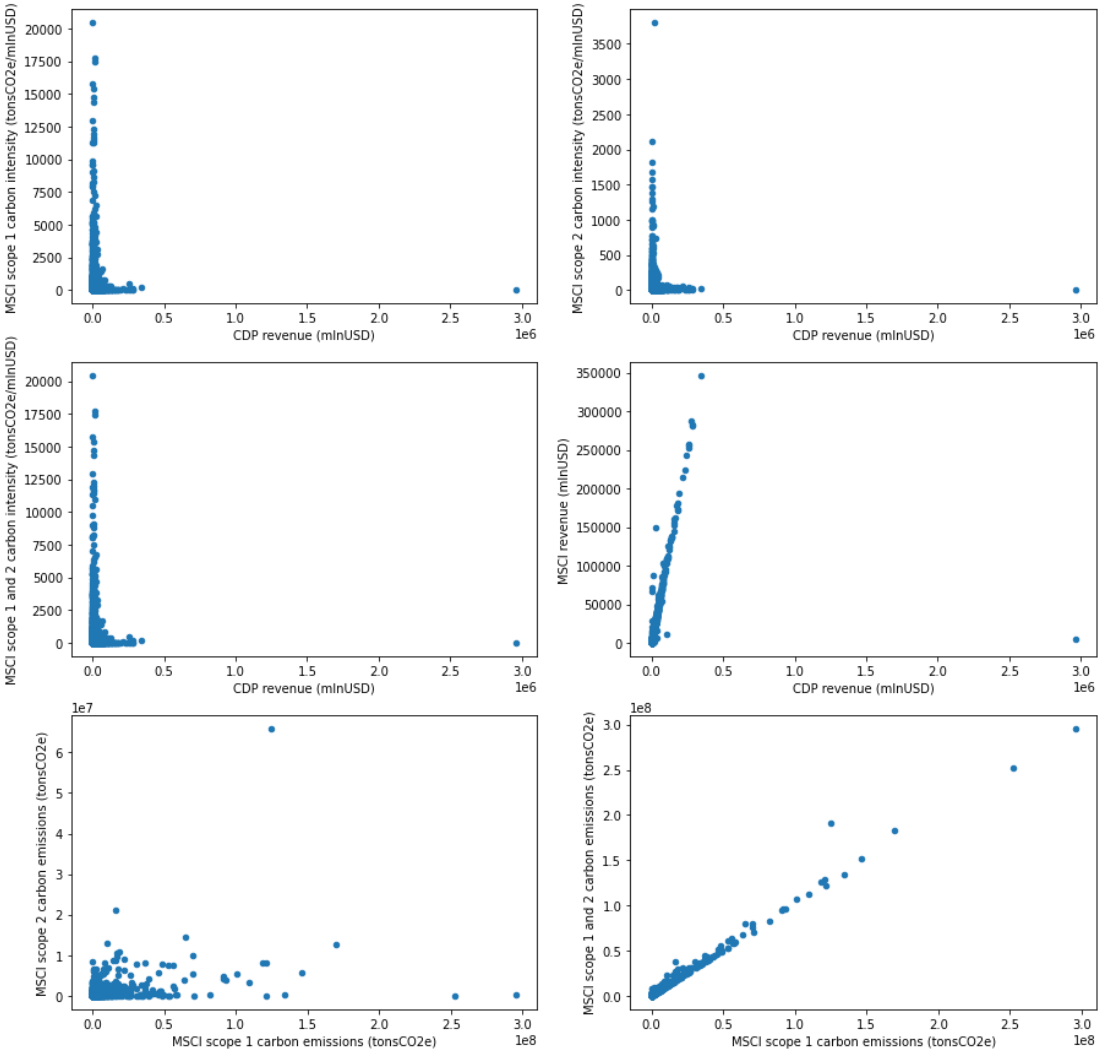

FIG. 58: The scatter plots between CDP revenue (mln USD) and other types of carbon emissions and carbon intensity, i.e., MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), MSCI scope 2 carbon intensity (tonsCO_2e/mln USD), MSCI scope 1 and 2 carbon intensity (tonsCO_2e/mln USD), and MSCI revenue (mln USD). The scatter plots between MSCI scope 1 carbon emissions (tonsCO_2e) and other types of carbon emissions and carbon intensity, i.e., MSCI scope 2 carbon emissions (tonsCO_2e), and MSCI scope 1 and 2 carbon emissions (tonsCO_2e).

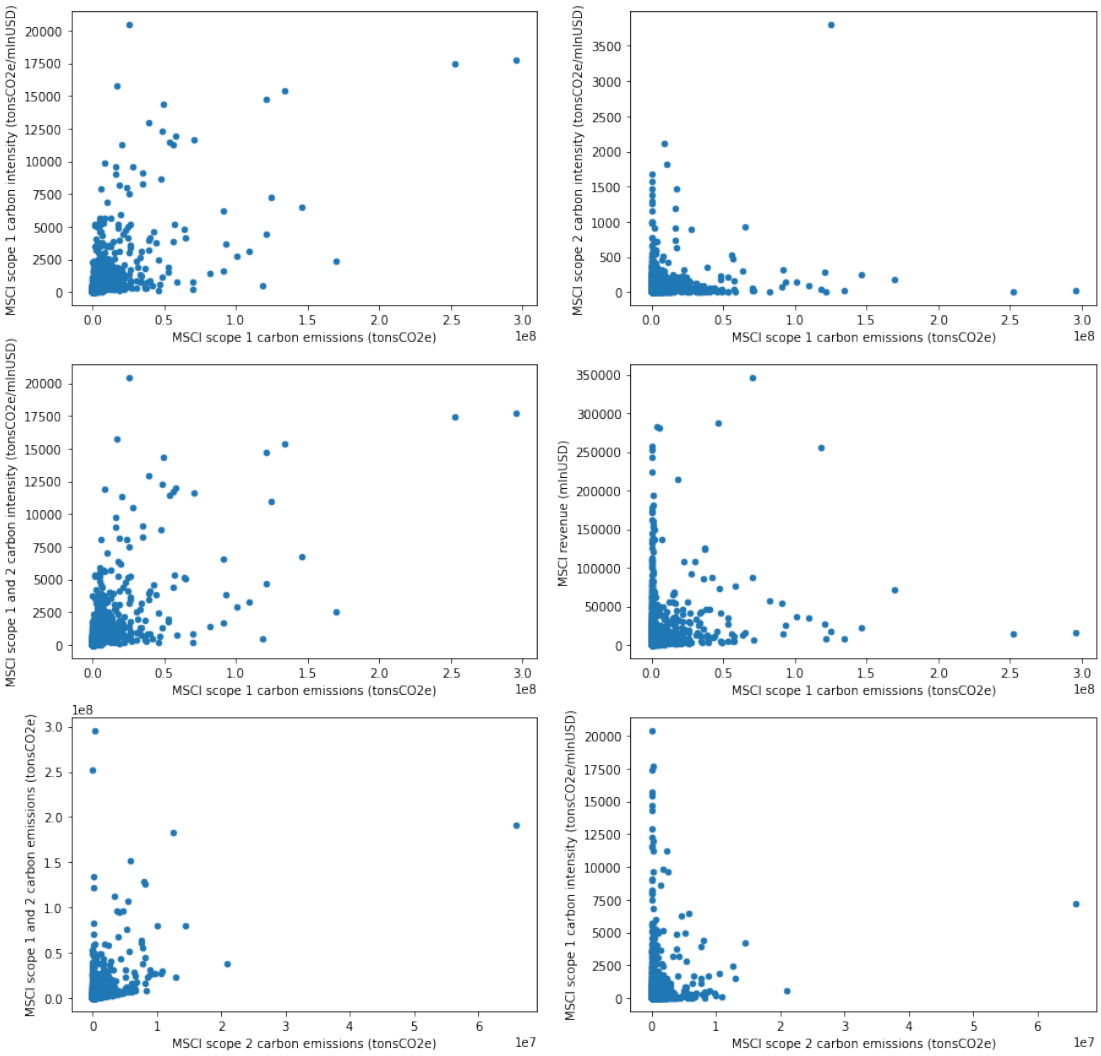

FIG. 59: The scatter plots between MSCI scope 1 carbon emissions (tonsCO_2e) and other types of carbon emissions and carbon intensity and the revenue, i.e., MSCI scope 1 carbon intensity (tonsCO_2e/mln USD), MSCI scope 2 carbon intensity (tonsCO_2e/mln USD), MSCI scope 1 and 2 carbon intensity (tonsCO_2e/mln USD), and MSCI revenue (mln USD). The scatter plots between MSCI scope 2 carbon emissions (tonsCO_2e) and other types of carbon emissions and carbon intensity, i.e., MSCI scope 1 and 2 carbon emissions (tonsCO_2e), and MSCI scope 1 carbon intensity (tonsCO_2e/mln USD).

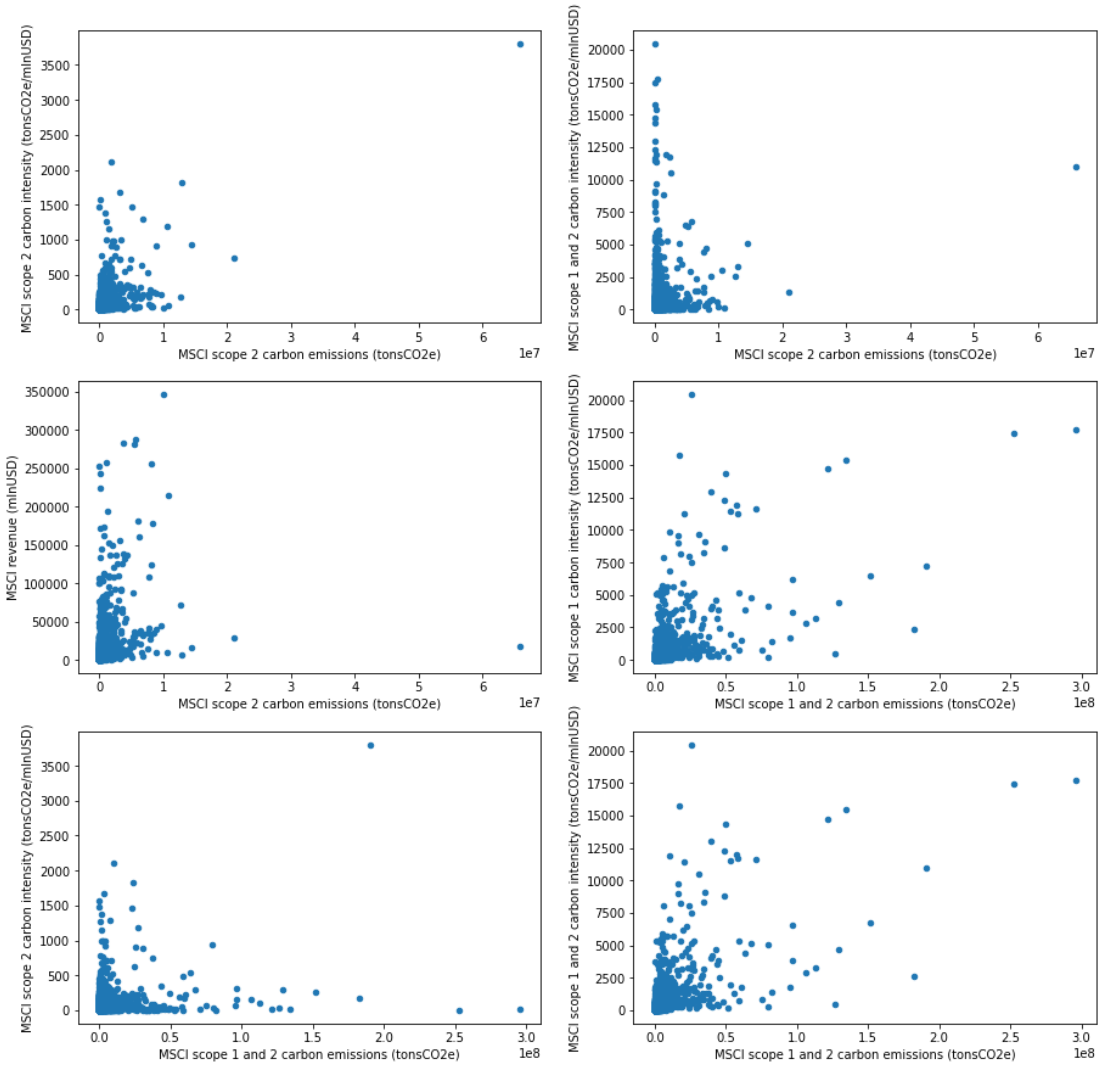

FIG. 60: The scatter plots between MSCI scope 2 carbon emissions (tonsCO_2e) and other types of carbon emissions and carbon intensity and the revenue, i.e., MSCI scope 2 carbon intensity (tonsCO_2e/mln USD), MSCI scope 1 and 2 carbon intensity (tonsCO_2e/mln USD), and MSCI revenue (mln USD). The scatter plots between MSCI scope 1 and 2 carbon emissions (tonsCO_2e) and other types of carbon emissions and carbon intensity, i.e., MSCI scope 2 carbon intensity (tonsCO_2e/mln USD), and MSCI scope 1 and 2 carbon intensity (tonsCO_2e/mln USD).

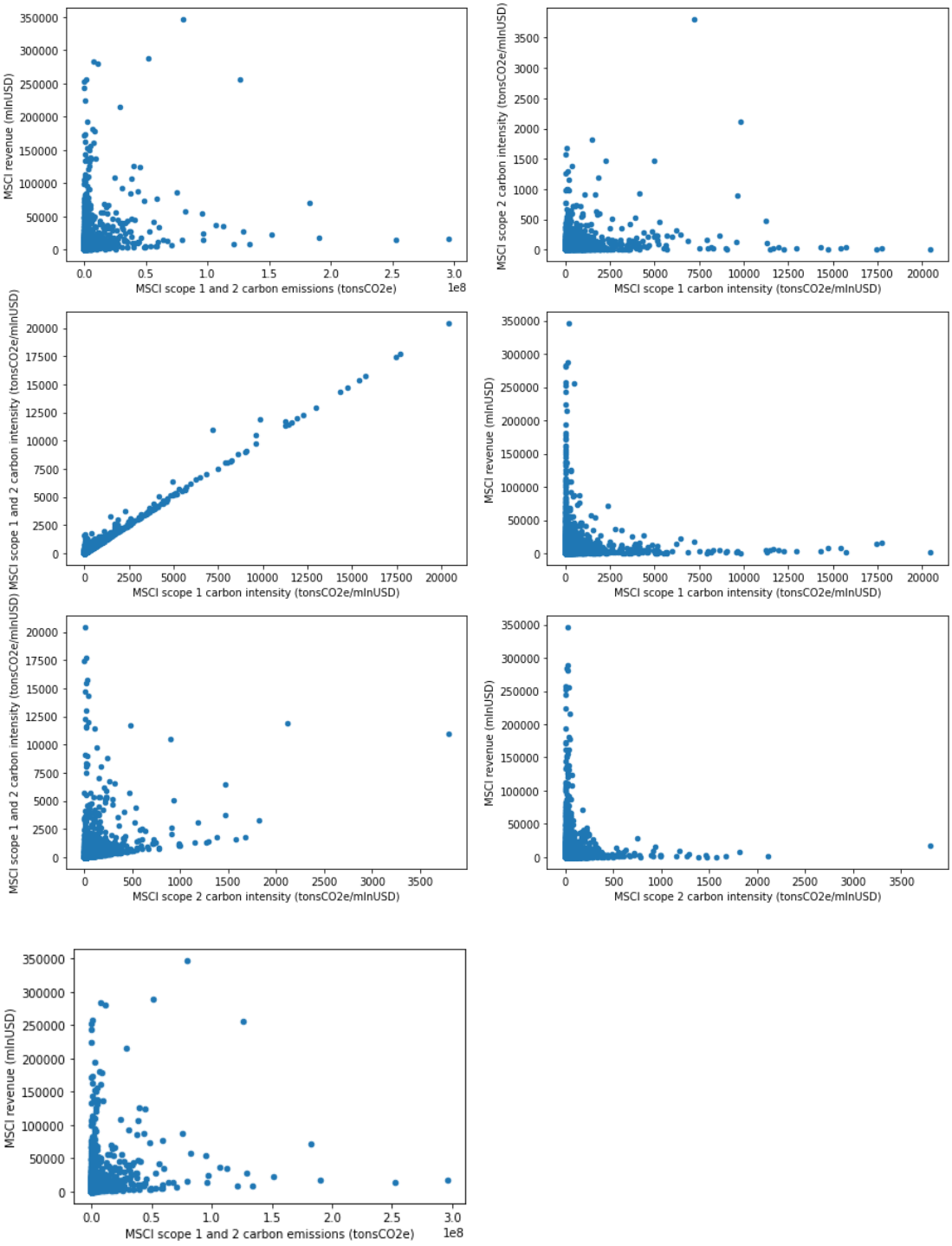

FIG. 61: The scatter plot between MSCI scope 1 and 2 carbon emissions (tonsCO_2e) and MSCI revenue (mln USD). The scatter plots between MSCI scope 1 carbon intensity (tonsCO_2e/mln USD) and other types of carbon emissions and carbon intensity and the revenue, i.e., MSCI scope 2 carbon intensity (tonsCO_2e/mln USD), and MSCI scope 1 and 2 carbon intensity (tonsCO_2e/mln USD), and MSCI revenue (mln USD). The scatter plots between MSCI scope 2 carbon intensity (tonsCO_2e/mln USD) and MSCI scope 1 and 2 carbon intensity (tonsCO_2e/mln USD), and MSCI revenue (mln USD). The scatter plot between MSCI scope 1 and 2 carbon intensity (tonsCO_2e/mln USD) and MSCI revenue (mln USD).