UNCERTAINTIES IN LCA

Variance-based global sensitivity analysis and beyond in life cycle assessment: an application to geothermal heating networks

Marc Jaxa-Rozen[1] · **Astu Sam Pratiwi**[1] · **Evelina Trutnevyte**[1]

Abstract

Purpose Global sensitivity analysis increasingly replaces manual sensitivity analysis in life cycle assessment (LCA). Variance-based global sensitivity analysis identifies influential uncertain model input parameters by estimating so-called Sobol indices that represent each parameter's contribution to the variance in model output. However, this technique can potentially be unreliable when analyzing non-normal model outputs, and it does not inform analysts about specific values of the model input or output that may be decision-relevant. We demonstrate three emerging methods that build on variance-based global sensitivity analysis and that can provide new insights on uncertainty in typical LCA applications that present non-normal output distributions, trade-offs between environmental impacts, and interactions between model inputs.

Methods To identify influential model inputs, trade-offs, and decision-relevant interactions, we implement techniques for distribution-based global sensitivity analysis (PAWN technique), spectral clustering, and scenario discovery (patient rule induction method: PRIM). We choose these techniques because they are applicable with generic Monte Carlo sampling and common LCA software. We compare these techniques with variance-based Sobol indices, using a previously published LCA case study of geothermal heating networks. We assess eight environmental impacts under uncertainty for three design alternatives, spanning different geothermal production temperatures and heating network configurations.

Results In the application case on geothermal heating networks, PAWN distribution-based sensitivity indices generally identify influential model parameters consistently with Sobol indices. However, some discrepancies highlight the potentially misleading interpretation of Sobol indices on the non-normal distributions obtained in our analysis, where variance may not meaningfully describe uncertainty. Spectral clustering highlights groups of model results that present different trade-offs between environmental impacts. Compared to second-order Sobol interaction indices, PRIM then provides more precise information regarding the combinations of input values associated with these different groups of calculated impacts. PAWN indices, spectral clustering, and PRIM have a computational advantage because they yield stable results at relatively small sample sizes ($n = 12,000$), unlike Sobol indices ($n = 100,000$ for second-order indices).

Conclusions We recommend adding these new techniques to global sensitivity analysis in LCA as they give more precise as well as additional insights on uncertainty regardless of the distribution of the model outputs. PAWN distribution-based global sensitivity analysis provides a computationally efficient assessment of input sensitivities as compared to variance-based global sensitivity analysis. The combination of clustering and scenario discovery enables analysts to precisely identify combinations of input parameters or uncertainties associated with different outcomes of environmental impacts.

Keywords Uncertainty · Sensitivity analysis · Clustering · Trade-off analysis · Scenario discovery

✉ Marc Jaxa-Rozen
   marc.jaxa-rozen@unige.ch

1   Renewable Energy Systems, Institute for Environmental
    Sciences (ISE), Section of Earth and Environmental
    Sciences, University of Geneva, 1205 Geneva, Switzerland

1 Introduction

Recent literature on life cycle assessment (LCA) reflects an increasing interest in methods for uncertainty and sensitivity analysis in order to understand how natural variability and epistemic uncertainties (i.e., incomplete knowledge) affect the environmental impacts of products and processes. Epistemic uncertainties in LCA can be

attributed to parameter, model, and scenario uncertainties (Rosenbaum et al. 2018). Parameter uncertainties for instance typically surround quantities used in inventory data or characterization factors (Huijbregts 1998). Model uncertainties may arise regarding the form of model equations or appropriate model boundaries (Reap et al. 2008). Scenario uncertainties relate to the use and interpretation of the model, such as the choice of a suitable functional unit or weighting method. Acknowledging these uncertainties is crucial to supporting decision-making: in comparative LCA, two products may yield impacts that are markedly different under nominal assumptions, but statistically indistinguishable after including uncertainties (Finnveden 2000). Conversely, specific uncertainties characterizing different products may propagate into unexpected outcomes and change the conclusions of the analysis. Methods for sensitivity analysis can support the interpretation of these uncertainties by identifying the model inputs and assumptions that are most influential toward model outputs. The ISO 14044 standard on LCA thus recommends including at least a sensitivity analysis in the evaluation and quality assurance of LCA results and underlines the contribution of methodological choices as well as input data toward uncertainty (International Organization for Standardization 2006).

Consistently with these recommendations, LCA publications and software environments increasingly incorporate methods for sensitivity analysis, systematically reviewed by Michiels and Geeraerd (2020). However, ISO 14044 does not recommend a specific technique, and thus, sensitivity analysis is often applied with a manual, "one-at-a-time" strategy, in which each input is varied separately to measure its impact on the LCA outputs. This strategy may fail to properly assess the influence of uncertain inputs in the presence of non-linearities or interactions between uncertain inputs (Saltelli and Annoni 2010). Automated methods for global sensitivity analysis (GSA) can address these issues by evaluating all uncertain model inputs and their impact on the LCA outputs at the same time, capturing the combined effects of model inputs by computing the model over multiple sampled values of these inputs. Applications of GSA in the LCA field (Wei et al. 2015; Bisinella et al. 2016; Lacirignola et al. 2017; Groen et al. 2017) have largely relied on variance-based GSA (Sobol 2001; Saltelli et al. 2004). A typical variance-based GSA identifies influential model inputs by computing first-order and total sensitivity indices (or so-called Sobol indices; Online Resource 1, Sect. 1). The first-order index is the fraction of the variance of the model's output that is caused by the uncertainty of each input on its own. The total index reflects the input's total contribution to the output variance when accounting for interactions with other inputs (Saltelli et al. 2004), such as when inputs multiply each other. Indices identifying these specific interactions, such as second-order

indices for interactions between pairs of inputs, can be estimated at an added computational cost.

While variance-based GSA is considered as best practice for LCA models, it has certain limitations due to its specific assumptions that should be considered before applying it (Saltelli et al. 2004). In particular: (*i*) the expected contribution of an input to the variance of the LCA output is a suitable measure of sensitivity, and (*ii*) the variance of the LCA output distribution describes the uncertainty of the output and consequently the uncertainty facing a decision-maker. However, the second assumption may not always hold, for instance, if changes in the central tendency of the output distribution or in its tails are more meaningful for the decision-maker (Saltelli 2002, "Methods"). Borgonovo (2006) links this second assumption to the mean-variance model used in classical decision theory, in which non-normal output distributions impose specific requirements on the preferences of the decision-maker. Borgonovo (2006) uses an analytical case to show that variance and thus variance-based sensitivity indices can be a misleading measure of uncertainty on a highly skewed output distribution. Liu et al. (2005) similarly show that variance-based GSA indices can be unreliable for non-normal, highly skewed output distributions. Variance-based GSA is based on propagating all uncertain input distributions through the model and hence yields a previously unknown and not necessarily normal output distribution of the model outputs. In LCA, uncertain variables in life cycle inventories commonly follow a log-normal distribution (Qin and Suh 2017), meaning that the log-transformed variable is normally distributed. However, the relationship between life cycle inventories and midpoint or endpoint impacts is not necessarily linear (Groen et al. 2017) so that the resulting distribution of impacts cannot be presumed to be (log-) normal. Alternative distributions are commonly used for uncertain input parameters, such as triangular distributions (Cucurachi et al. 2016; Muller et al. 2018). These distributions would similarly be propagated into non-normally distributed LCA model outputs. In this situation, distribution-based techniques for GSA (Borgonovo 2007; Cucurachi et al. 2016) may offer a more robust alternative to variance-based GSA by describing sensitivity through changes in the output distribution rather than through variance alone. These techniques, for instance, capture changes in the central tendency or in the tails of the output distribution that could be relevant for the decision-maker.

As an example of this approach, the PAWN distribution-based GSA technique (Pianosi and Wagener 2015, 2018; the PAWN acronym is derived from the authors' names) estimates sensitivity indices for each input by systematically computing Kolmogorov–Smirnov statistics on the conditional cumulative distribution function of the output. This technique (detailed in Online Resource 1, Sect. 2) is applicable regardless of the output distribution and has the added

benefit of being computationally efficient. It can be used with relatively small sample sizes and with datasets sampled with a generic Monte Carlo method, whereas the practical estimation of sensitivity indices with variance-based GSA typically requires numerical estimators (e.g., Saltelli et al. 2010; Online Resource 1, Sect. 1) and specific sampling designs to reduce computational cost (Lo Piano et al. 2021). These sampling designs may be impossible to implement with commercial LCA software that only provide generic Monte Carlo sampling functionality. The computational cost also grows with the number of parameters and may be intractable for models with a large number of parameters, when reliable estimation of variance-based sensitivity indices may require hundreds of thousands of model evaluations (Nossent et al. 2011; Butler et al. 2014). The PAWN technique should thus be advantageous in many LCA applications, but no application of PAWN exists in the LCA domain.

Whether derived from variance-based or distribution-based methods, sensitivity indices in GSA only provide general insights into the importance of uncertainties. As these indices are generally computed over the entirety of the model output, they do not indicate whether certain inputs may, for instance, be particularly influential toward causing the model output to fall in a specific decision-relevant range. Conversely, the indices do not indicate whether the estimated overall influence of certain inputs is due to specific values of these inputs or due to combinations of values across multiple inputs. However, these aspects are important for LCA practitioners who may need to meet impact thresholds for performance or regulatory compliance (Vidal and Sánchez-Pantoja 2019; Mahbub et al. 2019), or who want to explore combinations of input uncertainties that affect the relative performance of design alternatives (Heijungs et al. 2019). The GSA literature presents factor mapping as a suitable approach for these goals by identifying inputs most strongly associated with a specified range of outputs (Saltelli et al. 2004). This range of outputs can, for instance, be defined with a performance threshold. Proposed factor mapping methods include regionalized sensitivity analysis (Hornberger and Spear 1981), in which a binary threshold criterion is used to split a sample of model outputs. For each uncertain input, the cumulative distribution function of the two output groups is examined to assess whether a specific range of the input is associated with a shift in the cumulative distribution function of each group. Still, this method lacks statistical power in typical applications and is less informative about the influence of inputs for cases that involve interactions between multiple inputs (Spear et al. 1994; Saltelli et al. 2004). As such, applications of factor mapping are less common in the GSA literature (Pianosi et al. 2016).

By contrast, recent literature on model-based decision-making under deep uncertainty (reviewed in Marchau et al. 2019) presents methods such as scenario discovery, which is functionally similar to a factor mapping analysis but is better suited to study input interactions, for instance, by identifying the combinations of uncertain input values under which a design alternative would meet a performance threshold (Groves and Lempert 2007). Published applications of scenario discovery (reviewed in Kwakkel and Haasnoot 2019) commonly rely on the Patient Rule Induction Method (PRIM; Friedman and Fisher 1999). PRIM uses an optimization procedure to identify ranges of the uncertain model inputs which lead to a range of the model output specified by the analyst as interesting, such as output values above a certain threshold. This technique can be applied regardless of the shape of input or output distributions. A parallel approach was developed by Spear et al. (1994) in the form of the tree-structured density estimation technique, using classification and regression trees (CART) to explore interactions associated with decision-relevant ranges of a model's output. In this work, we use PRIM for scenario discovery as it offers more easily interpretable results for problems involving multiple uncertain inputs (Lempert et al. 2008). A clustering step can also be added to scenario discovery to pre-process the ensemble of model outputs generated in an uncertainty analysis or GSA. This enables analysts to group outputs into different trade-off patterns across multiple performance indicators or to identify different modes of behavior in cases without predefined performance thresholds. For example, spectral clustering is a graph-partitioning approach (reviewed in von Luxburg (2007) and Online Resource 1, Sect. 3) that is suitable for this purpose. It provides robust performance for clustering across multiple independent variables (such as multiple environmental impacts computed in a single model evaluation) on arbitrary output distributions. The PRIM technique can then be applied to identify the different combinations of uncertain input values that lead to each cluster of outputs (Jaxa-Rozen et al. 2016; Moallemi et al. 2017; Steinmann 2020; details in Online Resource 1, Sect. 4). In LCA, this combined approach could support the interactive analysis of trade-offs between multiple environmental impacts under uncertainty (Payen et al. 2015; Prado-Lopez et al. 2016) and identify the combinations of uncertain parameters causing different trade-off patterns.

In this study, we therefore combine distribution-based GSA (PAWN technique of Pianosi and Wagener 2018), spectral clustering, and PRIM, and illustrate how these techniques can yield more precise as well as additional insights toward understanding uncertainty in typical applications of LCA with non-normal output distributions, trade-offs across multiple impacts, and interactions between uncertainties. We apply these techniques to a case study of geothermal heating networks (Pratiwi and Trutnevyte 2021) and analyze

three system design alternatives defined on the basis of geothermal production temperature and heating network design. We compute environmental impacts under uncertainty for eight indicators, analyzed in the previous study: cumulative energy demand, fine particulate matter emissions, fossil and mineral resource scarcity, global warming, land use, terrestrial acidification, and water consumption. For each design alternative of the geothermal system, we highlight a typical use case for scenario discovery. In order to evaluate the added value of the three applied techniques, we compare their results to a variance-based GSA by computing Sobol first-order, second-order, and total indices (Sobol 2001). These three techniques of distribution-based GSA, clustering, and PRIM do not impose restrictions on the required sampling design or input distributions, and thus, they can be used with generic Monte Carlo sampling or with specific sampling already used in variance-based GSA. As a result, these techniques add minimal computing costs to an existing uncertainty analysis or variance-based GSA. To further make the three approaches more accessible, we use documented open-source Python packages for all analyses and provide an interactive code notebook replicating the general workflow for our LCA case as well as idealized test cases (Jaxa-Rozen et al. 2021). We note that our workflow is based on OpenLCA software with external processing in Python. Equivalent analysis packages are available in the R language, and our analysis can be reproduced with other LCA software that supports the export of Monte Carlo analysis results. This includes Brightway2, Umberto LCA + with the *LiveLink* feature, and the *PhD* and *Developer* license levels for SimaPro.

The case study, analysis methodology, and computational implementation are described in "Methods." This is followed by results from each technique in "Results of global sensitivity analysis" and "Results of spectral clustering and scenario discovery." "Discussion and conclusions" discusses these results and concludes the paper with recommendations for practitioners, along with avenues for further research.

## 2 Methods

2.1 Case study of geothermal heating networks

We illustrate our approach using the case study of LCA of geothermal heating networks in the State of Geneva in Switzerland. This case study builds upon the recent work of Pratiwi and Trutnevyte (2021), who compared the environmental performance of six alternatives of geothermal heating and cooling networks, defined on the basis of well depths and network design, in terms of eight environmental impacts (Table 1). The choice of impacts was made in the previous study, and it is based on a review of most common impact indicators used in energy studies by Dorning et al. 2019). In the earlier study of Pratiwi and Trutnevyte (2021), a minimum-maximum bounding analysis indicated that these environmental impacts were driven by relatively complex relationships among design parameters, uncertainties on material and energy intensity, and subsurface properties. For instance, while higher geothermal production temperatures may improve energy performance, this may be offset in a non-linear manner by impacts associated with deeper drilling. In parallel, the relative environmental performance of different geothermal system alternatives hinged on interactions among the design parameters, such as geothermal flow rate and geothermal production temperature. Different combinations of well depths and network designs were also associated with distinct trade-offs across the eight impacts. For these reasons, this application of LCA was a particularly relevant starting point for GSA, spectral clustering, and scenario discovery.

We narrowed down the six original geothermal network alternatives of Pratiwi and Trutnevyte (2021) into three design alternatives for the present study (Fig. 1), covering a greater range of geothermal production temperature and flow rate in each network design: (*i*) a networked heat pump alternative supplying a network of connected, decentralized heat pumps from the depth of 10–2200 m, (*ii*) a central heat pump alternative supplying district heating from the depth

Table 1 Environmental impacts of geothermal heating networks that are assessed in this study

Impact	Description	Unit
Cumulative energy demand	Total life cycle energy use from all sources (including geothermal energy)	MWh
Fine particulate matter emissions	Emissions of particulate matter with a diameter below 2.5 μm	kg $PM_{2.5}$ eq
Fossil resource scarcity	Depletion of fossil fuels available for future generations	kg oil eq
Global warming	Greenhouse gas (GHG) emissions from all sources	kg CO_2 eq
Land use	Species loss due to the change in land use	m^2-year crop eq
Mineral resource scarcity	Depletion of mineral resources	kg Cu eq
Terrestrial acidification	Deposition of inorganic substances modifying soil acidity	kg SO_2 eq
Water consumption	Water consumed over the life cycle, excluding water returned to the water body	m^3
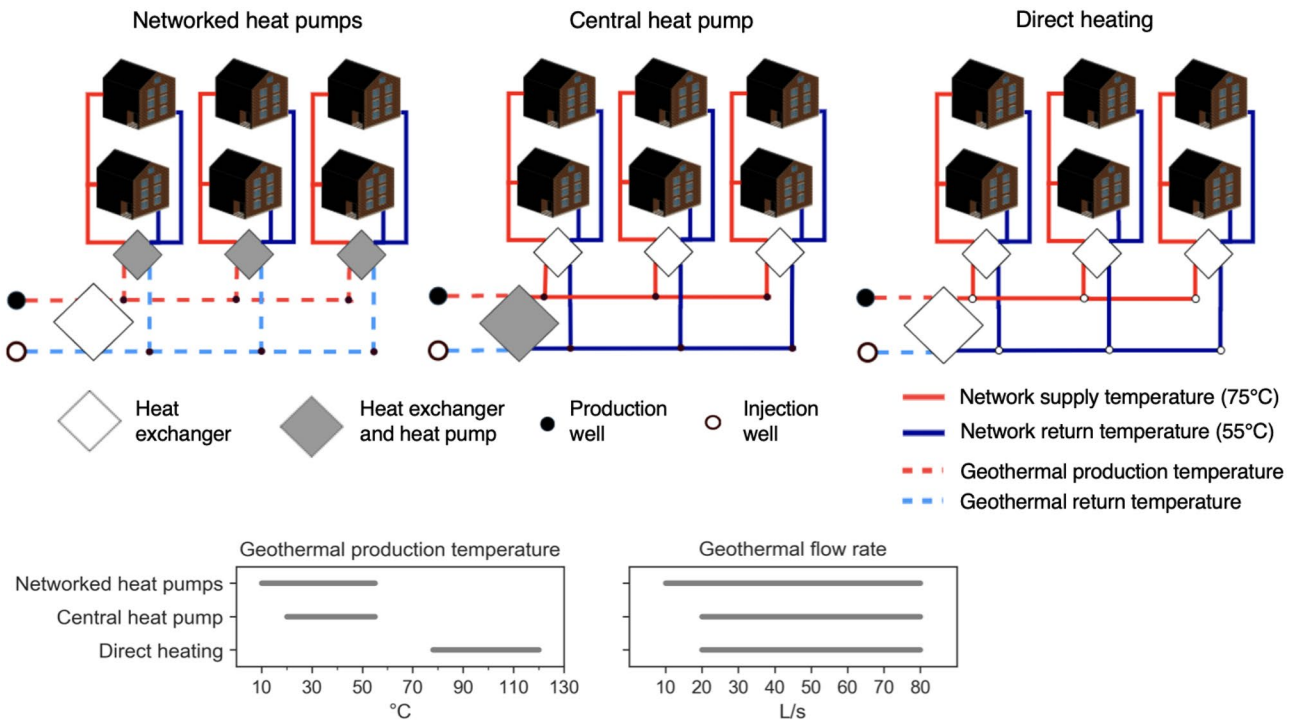
Fig. 1 Design alternatives for geothermal heating networks used in this study on the basis of the work by Pratiwi and Trutnevyte (2021). Top subplots: heating network design used in each alternative. Bot-

tom subplots: ranges of geothermal production temperature and flow rate in each alternative

of 300–2200 m, and (*iii*) a direct heating alternative supplying district heating from the depth of 2000–4400 m. In each alternative, the LCA boundaries include the drilling of geothermal wells, the construction of the heating network, the operation and maintenance of the network over its lifetime, and its decommissioning. We focus the analysis on heating and leave cooling out of scope. The results of Pratiwi and Trutnevyte (2021) indicated that operational electricity consumption contributed prominently to the life cycle impacts of the geothermal alternatives with heat pumps, using a hydropower-based electricity mix typical of conditions in the State of Geneva. For this study, we assume a more representative electricity supply mix based on Swiss national electricity grid (Itten et al. 2014). We otherwise maintain the other life cycle assessment choices documented by Pratiwi and Trutnevyte (2021), including the Ecoinvent 3.5 database (Wernet et al. 2016), OpenLCA 1.10 software, and ReCiPe 2016 H midpoint method (Huijbregts et al. 2017). Like the previous study, we report all impacts per MWh of delivered heat.

The original geothermal configurations contain between 41 and 43 uncertain input parameters (Online Resource 2, Table S1). As the sample size required for variance-based GSA increases with the number of parameters, using the original configurations for this analysis would be impractical due to the runtime of the OpenLCA implementation

(approximately 20 s per execution on an Intel Xeon 6126 CPU). To reduce the number of input parameters and enable a reference variance-based GSA, we therefore apply a preliminary screening step to identify non-influential parameters that can be removed from the analysis. To this end, we combine the ExtraTrees technique for non-linear regression (Geurts et al. 2006) with the PAWN distribution-based GSA technique. The Mean Decrease Impurity measure of input importance obtained from ExtraTrees approximates the relative value of Sobol total indices at small sample sizes (Jaxa-Rozen and Kwakkel 2018). It can therefore be used to assess the relative importance of model inputs toward output variance and can generally identify inputs that are non-influential in a full variance-based GSA. Nonetheless, the large initial number of parameters can make this estimation of input importances less reliable. To have a second independent assessment of input importances with a method that is applicable regardless of output distribution, we therefore use the PAWN technique that reliably ranks sensitivity indices for a non-normal distribution (Pianosi and Wagener 2018). We combine these methods to exclude input parameters that (*i*) do not meet a minimum threshold of relative importance using ExtraTrees, on any of the environmental impacts and any of the design alternatives, and (*ii*) do not meet a minimum threshold rank of importance using either ExtraTrees or PAWN, on any impact and any alternative.

This combined screening approach is visualized in Online Resource 2, Fig. S1.

We implement this preliminary screening step by generating separate ensembles of 12,000 input samples for each design alternative, using the `SALib` package (Herman and Usher 2017) to create a Latin hypercube sample (Online Resource 1, Sect. 2). We specify independent uniform distributions for each input parameter with the bounds reported in Online Resource 1, Table S1. We then use OpenLCA to compute impacts on each of the 12,000 samples and apply ExtraTrees regression using the scikit-learn Python package (Pedregosa et al. 2011) with the settings recommended in Jaxa-Rozen and Kwakkel (2018). In parallel, we use the `SAFE Toolbox` Python package (Noacco et al. 2019) to apply PAWN, with the same settings listed in "Computational implementation" for our detailed uncertainty analysis. We set the screening thresholds to exclude inputs that are not in the ten most influential parameters toward any impact, and that are not above 5% of the maximum relative importance on any impact using ExtraTrees. We find that relatively few inputs are influential across all impacts so that the set of retained inputs is robust to these screening thresholds. We thus identify 19 inputs in the networked and central heat pump alternatives (bolded in Table S1, Online Resource 2), and 18 in the direct heating alternative for which the heat pump coefficient of performance (COP) parameter is not used. To support the further analysis of the design alternatives with this reduced set of parameters using distribution-based GSA, clustering, and scenario discovery, we sample a new Latin hypercube ensemble of 12,000 samples for each alternative, only using the reduced set of influential parameters. To enable a meaningful comparison between design alternatives, we use consistent random seeds to initialize random number generation in Python, so that repeatable sequences of values are used in this sampling. A given uncertain parameter is therefore sampled using the same initial sequence in all design alternatives; the initial sequences are then rescaled to define input values based on the bounds specified in each alternative.

At this preparation step, we find that distributions of model outputs (i.e., environmental impacts) are consistently positively skewed with a longer upper tail (Fig. 2). To test whether the output distributions are normal, we use a Shapiro–Wilk test and find that the null hypothesis of a normal distribution
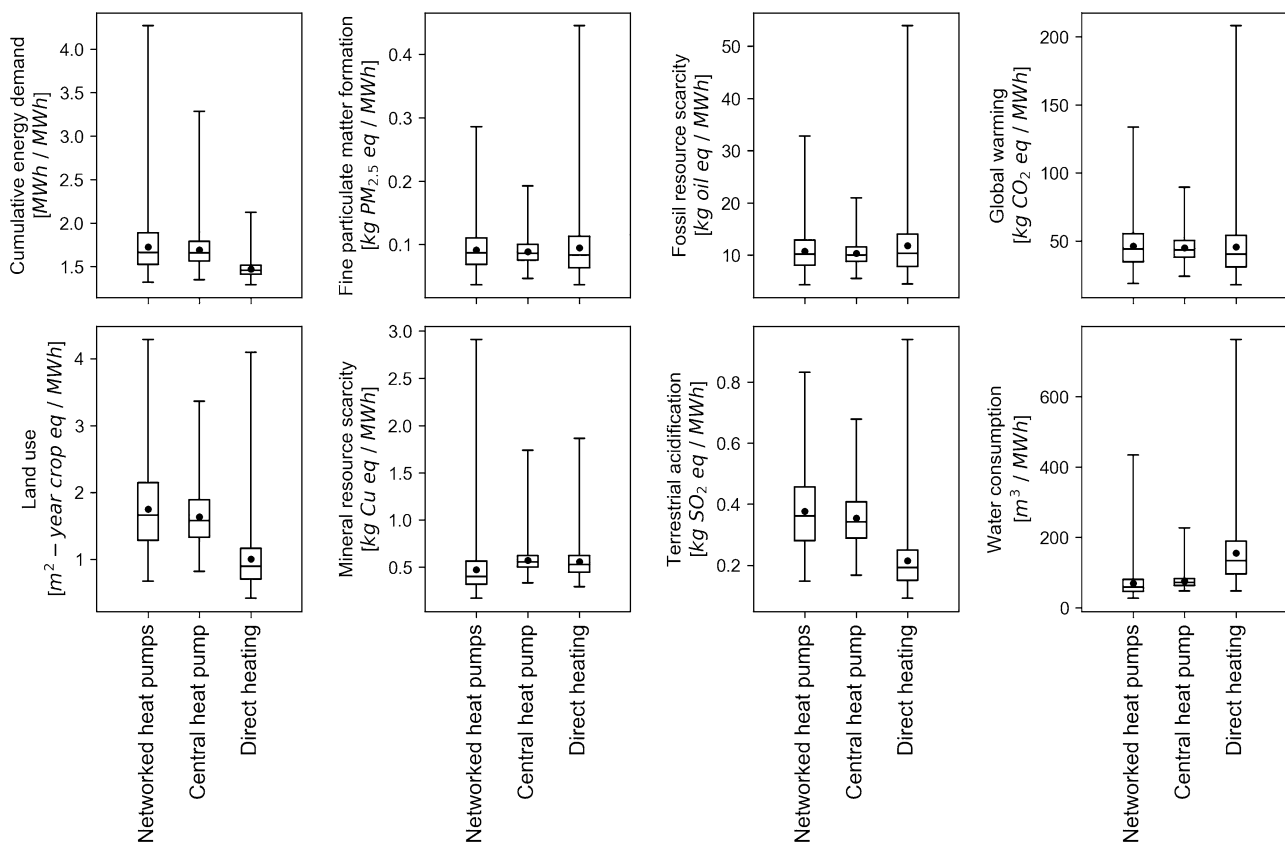
Fig. 2 Distribution of impacts in each design alternative, using Latin Hypercube ensembles ($n = 12,000$). All impacts are reported per MWh of delivered heat. For each boxplot, the mean is denoted by a black marker and the median by a black line. Boxes show the interquartile range; whiskers are extended to the full range of the data

is rejected at $p < 0.001$ for all impact distributions. We find the same result after log-transforming to test for a log-normal distribution. The skewness of the output distribution means that the expected contribution to variance described by the variance-based GSA indices may be unreliable as a metric of sensitivity (Liu et al. 2005; Borgonovo 2006). As numerical implementations of the Shapiro–Wilk test may be unreliable for large sample sizes (Razali and Wah 2011), we provide corresponding probability plots in Online Resource 2, Fig. S2 that show deviation from (log-)normality in the upper tail of values (using water consumption as an example). These results support the relevance of a distribution-based GSA approach in our case study. Figure S3 in Online Resource 2 directly visualizes the relative performance of design alternatives using a regret measure (Savage 1951); the direct heating alternative for instance has lower average impacts than the two heat pump alternatives on cumulative energy demand and land use, but consumes more water. Based on Pearson's r (Fig. S4, Online Resource 2), the impacts on fine particulate matter formation, fossil resource scarcity, and global warming are almost perfectly correlated in all design alternatives ($r > 0.99$). However, trade-offs may emerge between water consumption and the other impacts for the networked heat pump ($r < = 0.78$) and central heat pump ($r < = 0.58$) alternatives, justifying further trade-off analysis using clustering. The direct heating alternative yields highly correlated impacts ($r > 0.93$), except in relation to cumulative energy demand.

2.2 Methodology for the uncertainty analysis

We use the uncertainty screening as a starting point to systematically investigate input sensitivities, interactions, and trade-offs between environmental impacts. Using the reduced set of influential parameters identified by screening, we first perform variance-based and distribution-based GSA on each geothermal heating network alternative, to identify influential single inputs with Sobol first-order and total indices (Online Resource 1, Sect. 1) and PAWN indices (Online Resource 1, Sect. 2). In order to identify influential interactions between pairs of input parameters, we extend the variance-based GSA with Sobol second-order indices. We then combine clustering and scenario discovery to understand how these influential inputs and interactions lead to

different subsets of scenarios of interest, that present different trade-offs between impacts. We define a *scenario* as the combination of the vector of input samples used in a given execution of the model, with the resulting vector of environmental impacts. In principle, the subsets of interest for scenario discovery could be defined through a targeted approach by setting performance thresholds on one or multiple model outputs, or they could also be defined from an exploratory perspective by applying statistical learning techniques to find patterns in the model outputs. We combine these approaches to illustrate typical use cases at different levels of complexity. As such, we first choose a representative performance measure to define subsets of interest in each design alternative (Table 2): an impact threshold, a threshold for the relative regret between alternatives, and a multi-objective measure of relative performance based on Pareto dominance (Ravalico et al. 2009), where an alternative A is Pareto-dominated by an alternative B if it is no better than B in all objectives (eight environmental impacts), and worse in at least one objective.

We use water consumption to define an example of an impact threshold and a relative regret threshold; this impact is broadly distributed (Fig. 2), and it was found by Pratiwi and Trutnevyte (2021) to be driven by multiple uncertain life-cycle activities for geothermal heating, making it a particularly relevant case to illustrate our analysis. For each subset of interest, we then (*i*) apply spectral clustering (Online Resource 1, Sect. 3) in order to identify different patterns of trade-offs between impacts that emerge within the subset, and (*ii*) apply the PRIM scenario discovery technique (Online Resource 1, Sect. 4) in order to identify the combinations and ranges of uncertain model inputs associated with each cluster (i.e., each pattern of trade-offs). Figure 3 summarizes the relationships between these analysis techniques. We visualize the results for the networked heat pump alternative in the main text and provide results for the two other alternatives in Online Resource 2.

2.3 Computational implementation

We implement techniques for global sensitivity analysis, spectral clustering, and scenario discovery using open-source packages in the Python 3.x scientific computing

Table 2 Definition of subsets of scenarios of interest for each design alternative

Design alternative	Type of performance measure	Definition of the subset of scenarios of interest
Networked heat pumps	Single-objective impact threshold	Scenarios with water consumption < 40 m^3/MWh ($n = 1200$)
Central heat pump	Single-objective relative performance	Scenarios with water consumption regret $> 50\%$ relative to networked heat pump alternative ($n = 1,743$)
Direct heating	Multi-objective relative performance	Scenarios that are Pareto-dominated by central heat pump alternative ($n = 391$)
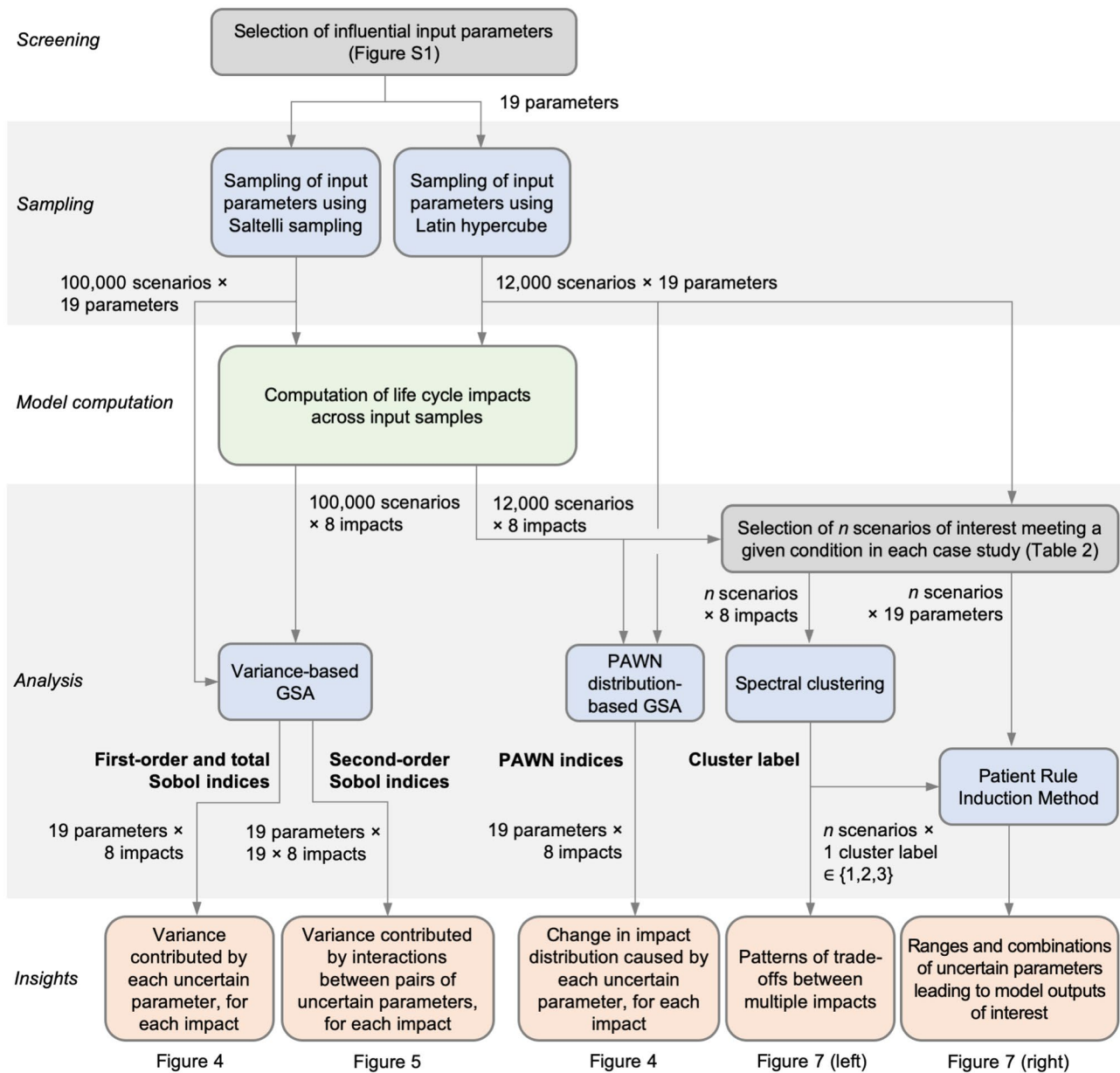
Fig. 3 Overall structure of the analysis applied for each design alternative. Online Resource 1 provides theoretical background for each method in the sampling and analysis levels. Fig. S1, Online Resource 2 details the preliminary screening step used to reduce the number of model parameters

ecosystem. Equivalent implementations are available in R, for instance using the sensitivity and OpenMORDM packages (Hadka et al. 2015; Pujol et al. 2017). We implement a Python parallelization wrapper based on OpenLCA's interprocess communication (IPC) interface to distribute the model executions across 16 processing cores. Python code for this OpenLCA interface and our overall implementation is documented by Jaxa-Rozen et al. (2021).

To define input values for the variance-based GSA, we use the quasi-random Sobol sampling sequences recommended by Saltelli et al. (2010), as implemented in the

SALib package (Herman and Usher 2017). To estimate first-order (S1), second-order (S2), and total (ST) Sobol indices, these sequences require the model to be computed for a total of $N = n(2k + 2)$ input samples, where n is a baseline sample size and k is the number of input parameters. We choose $n = 2500$ to improve the convergence of Sobol indices within computational constraints, for a total of $N = 100,000$ samples for each design alternative. In order to estimate the value of the Sobol indices, we first compute environmental impacts on each input sample with OpenLCA, then use `SALib` to apply the estimators of Saltelli

et al. (2010), averaged over 100 bootstrap resamples for each of the eight environmental impacts. The sampling and estimators are detailed in Online Resource 1, Sect. 1. The ranking and relative importance of total indices are both largely stable for $N > 50,000$ scenarios (Online Resource 2, Fig. S5). The convergence of second-order indices is less satisfactory (Online Resource 2, Fig. S6); larger sample sizes were limited by memory restrictions. We then compute distribution-based PAWN indices for each design alternative, using the SAFE Toolbox package (Noacco et al. 2019). To estimate the conditional cumulative distribution functions for each uncertain input, we divide the range of each input with the default setting of 10 conditioning intervals used by the package. We then use the median value of the Kolmogorov–Smirnov statistic across these conditioning intervals to define the PAWN median index of each input (averaging over 100 bootstrap resamples). We repeat this process for each environmental impact. We find that the rankings and relative importance of the PAWN indices are stable for $N > 10,000$ samples. To make the analysis computationally more efficient, we therefore compute these indices on the smaller ensembles of 12,000 Latin hypercube samples generated after the screening step, rather than the larger ensembles used for variance-based GSA. To check for extreme values (Pianosi and Wagener 2015), we also compute alternative indices using the maximum value of the Kolmogorov–Smirnov statistic across the same conditioning intervals.

In order to analyze patterns of trade-offs between environmental impacts, we use the spectral clustering module included in scikit-learn after normalizing all impacts to [0,1] within the subset of scenarios of interest selected for each design alternative. We select these scenarios of interest from the same ensembles of 12,000 Latin hypercube samples that were used to compute PAWN indices. We consider the eight environmental impacts calculated for each sample as independent dimensions on which the scenarios should be clustered. We select the default radial-basis function kernel to weight the graph representation of these samples on the basis of the similarity of their environmental impacts (Online Resource 1, Sect. 3) and test a range of two to six clusters in each design alternative. We use the scikit-learn grid search function to test a range of 0.001 to 1 for the γ kernel parameter. Based on visual evaluation and a silhouette metric (Rousseeuw 1987), we choose three clusters for all design alternatives. A greater number of clusters tend to divide existing clusters rather than highlighting new trade-off patterns. The γ parameter is set to 0.001, 0.3, and 0.1 for the networked heat pump, central heat pump, and direct heating alternatives. These parameters need to be adjusted to each specific case depending on the properties of the underlying data, such as its distribution and the number of dimensions (impacts) used for clustering, as

few theoretical guidelines exist (von Luxburg 2007). For each design alternative, this clustering step yields a vector of 12,000 integers, corresponding to the size of the original model outputs computed over the Latin hypercube samples; the original model output values are replaced with an integer value representing the cluster label (1, 2, or 3).

For each design alternative, we then apply the PRIM technique (Friedman and Fisher 1999) on the vector of labels produced by clustering, to identify combinations and ranges of input parameters associated with each of the clusters. This technique assumes that model outputs of interest (i.e., outputs with a specified cluster label) are associated with a (hyper-)rectangular region (or "box") of the input space, after sequentially removing ranges of the input that do not lead to these outputs of interest (Online Resource 1, Sect. 4). However, this is not always the case (Dalal et al. 2013), and the quality of the identified box of input ranges must be assessed for each application. To this end, the scenario discovery literature typically uses the trade-off between *coverage* and *density* to assess the performance of the algorithm (Lempert et al. 2008), where coverage is the fraction of all scenarios of interest that fall within the box identified by PRIM, and density is the fraction of scenarios within the box that is of interest. We use the implementation of PRIM included in the EM Workbench package (Kwakkel 2017), which interactively visualizes the trade-off between coverage and density and the revised objective function introduced in Kwakkel and Jaxa-Rozen (2016) to improve the analysis of categorical parameters. We visualize boxes meeting an arbitrary coverage threshold of 75%, so that the combinations of ranges of input parameters identified by PRIM describe at least this fraction of the total scenarios in each cluster.

3 Results of global sensitivity analysis

3.1 Single-input sensitivity indices

In the analyzed case study, the variance-based and distribution-based indices generally provide consistent estimations of the relative importance of uncertain inputs (Fig. 4). The geothermal production temperature is largely dominant across the environmental impacts and sets of sensitivity indices with the exception of the impact on water consumption, where geothermal flow rate is more influential. The rank correlations between Sobol total indices and PAWN median indices are relatively high ($\rho \geq 0.6$; Online Resource 2, Fig. S7), except in the case of terrestrial acidification. Taking rank correlations across impacts for the Sobol total indices alone, correlations are particularly strong ($\rho > 0.8$) between these indices for fine particulate matter formation, fossil resource scarcity, global warming, land use, and terrestrial acidification, indicating that these impacts are largely driven by the same parameters. This
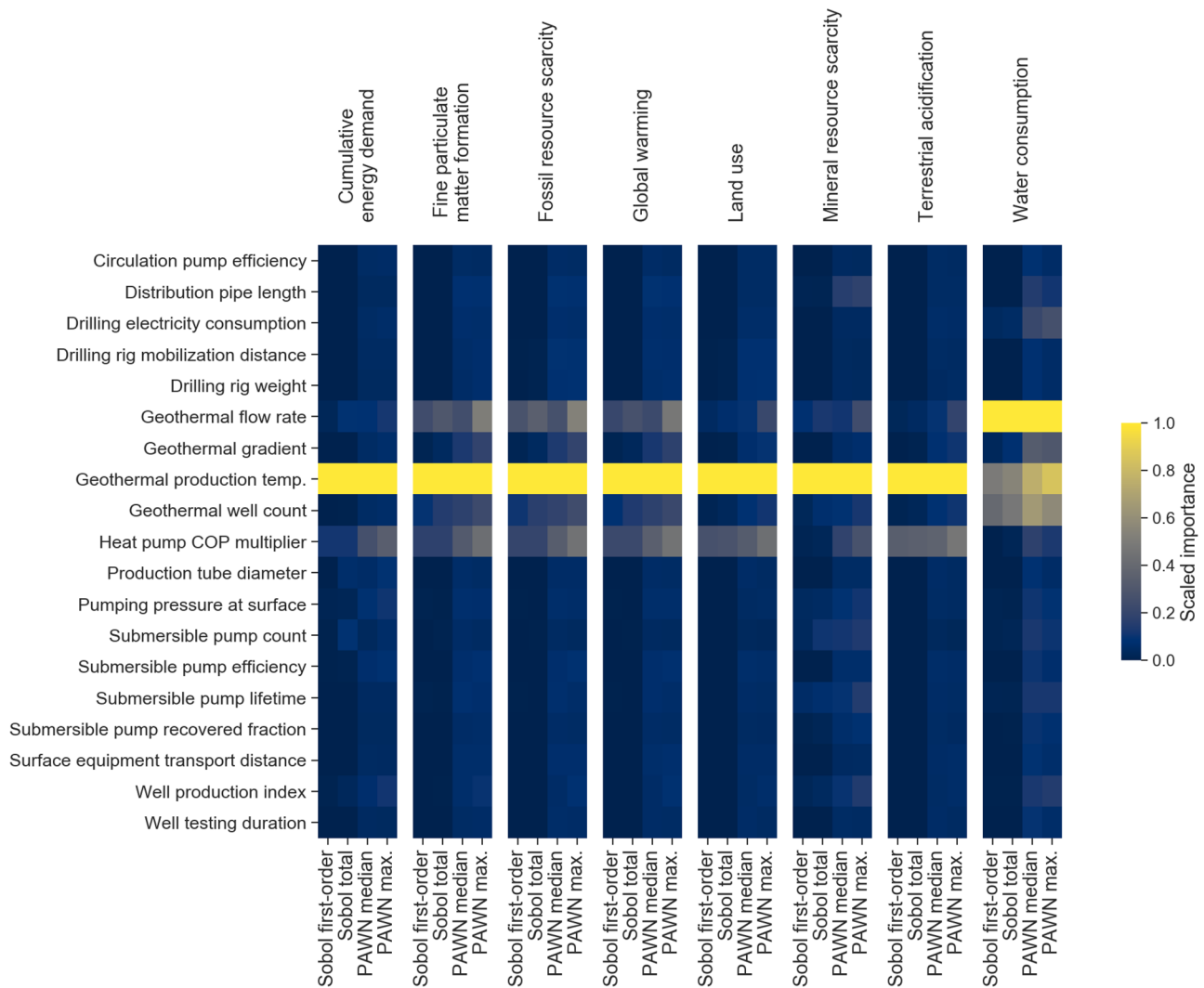
Fig. 4 Relative variable importances for the networked heat pump design alternative, grouped by environmental impact (subplots) and set of sensitivity indices (subplot columns). For each environmental impact, the vector of sensitivity indices is normalized to [0,1] in each set of sensitivity indices separately. The relative variable importances for other alternatives are available in Online Resource 2, Fig. S10 and Fig. S13

finding is consistent with Pearson's *r* for correlations between the impacts (Online Resource 2, Fig. S4). Cross-impact rank correlations are weaker for the PAWN median indices, indicating greater variation in the rankings, but they remain strong ($\rho > 0.8$) between PAWN indices for fine particulate matter formation, fossil resource scarcity, and global warming. Between the Sobol first-order and total indices, relative importances are generally consistent. However, the higher absolute value of total indices in several cases (Online Resource 2, Fig. S8) denotes higher-order interactions, for example, involving geothermal flow rate, geothermal production temperature, and number of wells for the impact on water consumption. Between the distribution-based PAWN median and maximum indices, relative values are also largely consistent, but the maximum indices for geothermal flow rate are higher on fine particulate

matter formation, fossil resource scarcity, and global warming. The heat pump COP multiplier presents a similar trend on most impacts. This indicates that a limited region of the input range for these parameters is associated with a particularly significant impact on the output cumulative distribution function (Pianosi and Wagener 2015) and may need to be investigated further.

Despite the relatively high correlation between Sobol total indices and PAWN rankings, some of the PAWN relative importances are consistently higher than the variance-based indices, particularly in the case of the heat pump COP multiplier and the geothermal gradient. Following the definition of Saltelli et al. (2004), the first-order index is the reduction in variance which could be expected if the parameter was fixed. However, in the case of water consumption as an example, the geothermal gradient

(S1 = 0.02; PAWN median = 0.29) yields a conditional variance that is larger than unconditional variance within a large portion of its uncertainty range (Online Resource 2, Fig. S9). Fixing it to its lower bound would thus increase the unconditional variance of water consumption, i.e., the uncertainty of the output would counterintuitively be increased by resolving the uncertainty of this parameter. While this effect is not exceptional on skewed output distributions and when interactions between variables are significant (Borgonovo 2006; Saltelli et al. 2004, Box 2.4), this result underlines that the expected reduction in variance—which is, by definition, taken over the entire uncertainty range—may not always be the most meaningful measure of sensitivity. As such, the PAWN indices here may better highlight the geothermal gradient's impact on water consumption.

In the case of the central heat pump design alternative, we similarly find that relative importances are generally consistent across the different sensitivity indices (Online Resource 2, Fig. S10; rank correlations in Fig. S11). Relative importances are dominated by the production temperature and heat pump COP multiplier for most impacts except water consumption, on which geothermal flow rate and the number of geothermal wells are most influential. Compared to the networked heat pump alternative, a greater variety of parameters is influential across impacts as the narrower range of production temperatures used in the central alternative makes this parameter less salient. Across all indices, parameters related to the submersible well pumps and distribution piping thus have more influence on mineral resource scarcity than with the networked heat pump design alternative. Absolute differences between Sobol first-order and total values (Online Resource 2, Fig. S12) indicate the presence of interactions, for instance, involving geothermal production temperature in the case of mineral resource scarcity. The relative importances of PAWN median and maximum indices are largely consistent, so that the outputs do not appear particularly sensitive to specific regions of the input ranges. However, both sets of PAWN indices present some discrepancies with the variance-based indices. Using the former, the geothermal flow rate and geothermal gradient are relatively more influential on fine particulate matter formation, fossil resource scarcity, and global warming. Conversely, the flow rate is relatively less influential on water consumption using the PAWN indices than with the variance-based indices. The long upper tail of this outcome distribution may make variance less reliable as a description of uncertainty so that this impact could be further explored with a visualization of conditional variance across the range of flow rate, as in the case of the geothermal gradient for the networked heat pump alternative.

For the direct heating design alternative (Online Resource 2, Fig. S13; rank correlations in Fig. S14), the number of wells and geothermal flow rate dominate relative importances except for the case of cumulative energy demand, where production temperature ranks ahead of flow rate. Sobol total index rankings are almost perfectly correlated ($\rho > 0.98$) across the impacts of fine particulate matter formation, fossil resource scarcity, global warming, land use, and terrestrial acidification. However, they are less correlated with the PAWN rankings than in the shallower alternatives ($\rho < 0.82$). The consistent difference between Sobol first-order and total indices for flow rate and number of wells (Online Resource 2, Fig. S15) indicates a likely interaction between these parameters. The relative difference between PAWN median and maximum indices indicates that particular regions of the input ranges may be especially sensitive in the case of production temperature (on cumulative energy demand) and well count (on water consumption). Overall, as the direct heating alternative does not use a heat pump, other parameters associated with the operation phase (i.e., efficiencies of the circulation pump and submersible well pump, and pumping pressure at the surface) emerge as more influential than in the heat pump alternatives, particularly for cumulative energy demand. This effect is especially notable with the PAWN indices. Similarly, the geothermal gradient is slightly but consistently more influential with the PAWN indices than with the variance-based indices, likely indicating a similar effect on conditional variance as illustrated for the networked heat pump design alternative.

3.2 Pairwise interaction indices

The differences found between first-order and total indices justify the investigation of pairwise interactions using second-order Sobol indices (Fig. 5; other impacts, along with absolute index values, are shown in Online Resource 2, Fig. S16 due to their high correlation with impacts shown in the main text). For the networked heat pump design alternative, the variance of cumulative energy demand is dominated by first-order indices. However, consistently with differences between first-order and total indices (Online Resource 2, Fig. S8), we find more substantial interactions toward impacts on global warming (mainly between geothermal flow rate and the number of wells) and mineral resource scarcity (between geothermal production temperature and the number of wells, or submersible pumps and the well production index). In relation to total variance, the largest interactions occur in the context of water consumption. Taking the most influential parameters for this impact, pairwise interactions thus account for most of the variance contributed by the geothermal flow rate and by the number of wells, and for nearly all the variance contributed by the geothermal gradient. Pairwise interactions with these three parameters also represent a large part of the production temperature's contribution to water consumption variance. These parameters all play a key role
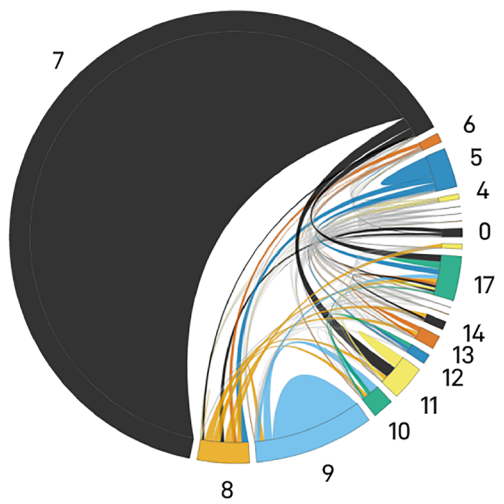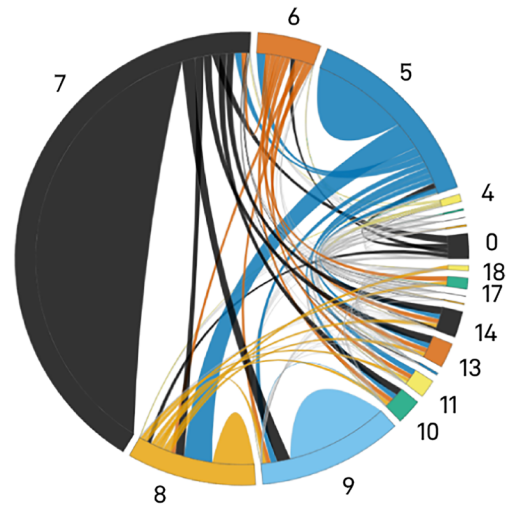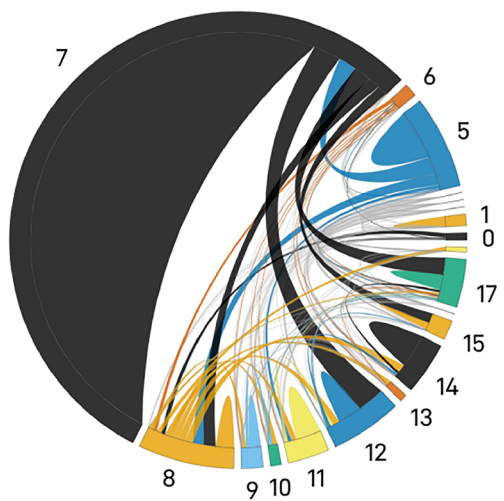
0: Circulation pump efficiency
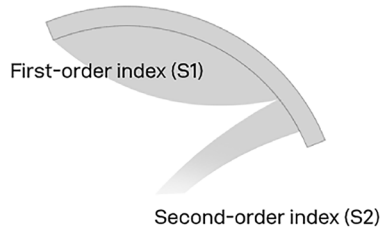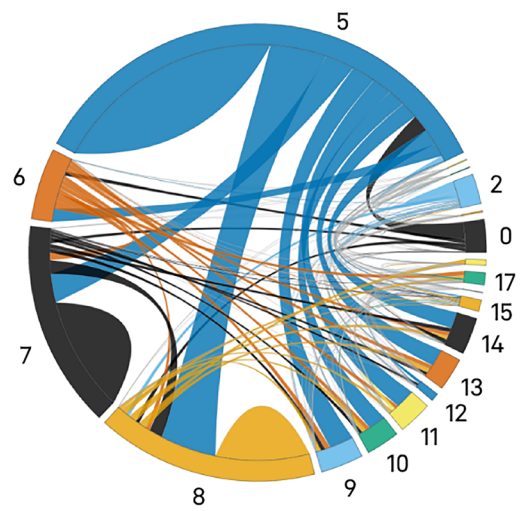1: Distribution pipe length
2: Drilling electricity consumption
3: Drilling rig mobilization distance
4: Drilling rig weight
5: Geothermal flow rate
6: Geothermal gradient
7: Geothermal production temperature
8: Geothermal well count
9: Heat pump COP multiplier
10: Production tube diameter
11: Pumping pressure at surface
12: Submersible pump count
13: Submersible pump efficiency
14: Submersible pump lifetime
15: Submersible pump recovered fraction
16: Surface equipment transport distance
17: Well production index
18: Well testing duration

◄**Fig. 5** Sobol first-order and second-order indices for the networked heat pump design alternative; additional impacts are visualized in Online Resource 2, Fig. S16. For each chord diagram, the total circumference represents the variance of each impact. The arc length of each portion of the circumference represents the relative contribution of each input to the variance (total index). Within this portion, the arc length of a self-referential chord represents the first-order index; second-order indices are represented by the arc length (thickness) of chords linking two inputs

for the drilling of geothermal wells, which was found to significantly affect water consumption in the earlier analysis by Pratiwi and Trutnevyte (2021). However, we also emphasize that the values of these second-order indices should be interpreted with caution due to the limited convergence observed at tractable sample sizes (Online Resource 2, Fig. S6).

In the case of the central heat pump design alternative (Online Resource 2, Fig. S17), we similarly find that the most influential pairwise interactions arise in the case of mineral resource scarcity (between geothermal production temperature, or the number of wells and submersible pumps) and water consumption (particularly between geothermal flow rate and the number of wells, whereas the geothermal gradient is involved in multiple weaker interactions with the drilling-related parameters). This is consistent with the differences observed between first-order and total indices on these impacts. Finally, the direct heating alternative (Online Resource 2, Fig. S18) presents the most salient interactions: the geothermal flow rate and the number of wells dominate pairwise interactions across all impacts, consistently with differences between first-order and total indices, and present the highest absolute second-order index values of any of the configurations: 7–10% of output variance depending on the impact. As with the shallower alternatives, the geothermal gradient is also involved in weaker interactions with multiple variables related to the drilling of geothermal wells.

## 4 Results of spectral clustering and scenario discovery

In the case of the networked heat pump design alternative, the subset of selected scenarios for scenario discovery, which is defined using a maximum water consumption threshold of 40 m^3/MWh, yields a relatively broader range of outcomes across the other impacts (Fig. 6). Consistently with cross-impact correlation coefficients (Online Resource 2, Fig. S4), a low water consumption impact may for instance be accompanied by higher impacts on terrestrial acidification. We decompose this subset using spectral clustering (Fig. 7, left column), finding three clusters which rather evenly span the range of impacts found in the subset. Cluster 1 is thus a "best-case" group of scenarios across all impacts, while cluster 2 is, in comparison, a "worst-case" group, including

several high outliers on cumulative energy demand. Cluster 3 displays a different trade-off pattern, sharing cluster 1's low impacts except for mineral resource scarcity and water consumption. As expected from the impact correlation coefficients, impacts within each cluster are similar across fine particulate matter formation, fossil resource scarcity, global warming, land use, and terrestrial acidification.

The significant uncertainties identified by PRIM (Fig. 7, right panel) are consistent with the GSA results: the geothermal flow rate, geothermal production temperature, and number of wells are the only variables that appear significant in all three clusters. These were previously identified as the three most influential variables on water consumption, based on both types of single-variable sensitivity indices and Sobol second-order indices. Considering all clusters, we can conclude that at least 75% of scenarios with a water consumption below 40 m^3/MWh in this alternative have a relatively high geothermal fluid flow (≥ 34 L/s), combined with a relatively high production temperature (≥ 24 °C), and a small number of wells (≤ 4). However, within these ranges, different patterns of outcomes emerge from different input combinations. The relatively worse outcomes of cluster 2 are thus associated with a medium range of production temperatures (24–45 °C). Clusters 1 and 3 include similar combinations of flow rate, production temperature, and heat pump COP multiplier. However, relaxing the constraint on the number of wells and removing the constraint on geothermal gradient leads to higher impacts on water consumption in cluster 3. The presence of the geothermal gradient as a significant variable in cluster 1 is coherent with the PAWN indices, which highlighted this variable's impact on water consumption more prominently than the variance-based indices. As such, the best-performing scenarios on water consumption occur with a relatively high geothermal gradient (≥ 0.025 °C/m), combined with favorable ranges of the other significant parameters. Whereas the second-order interaction indices offered relatively poor convergence even for $n = 100{,}000$ scenarios, we note that PRIM can here identify meaningful parameter combinations with a much smaller sample.

In the case of the central heat pump alternative, we select scenarios for which water consumption is more than 50% higher than in the networked heat pump alternative (Online Resource 2, Fig. S19). In relation to the full ensemble of central heat pump scenarios, we find that the scenarios of interest nonetheless occur in the context of low impacts on water consumption. The clustering step identifies three clusters (Online Resource 2, Fig. S20). Cluster 1 has the lowest outcomes across all impacts, while cluster 2 identifies a smaller group including the highest outcomes on all impacts except for water consumption. Cluster 3 contains the majority of the scenarios with midrange values on most impacts, but a broad distribution of water consumption. The PRIM input ranges indicate
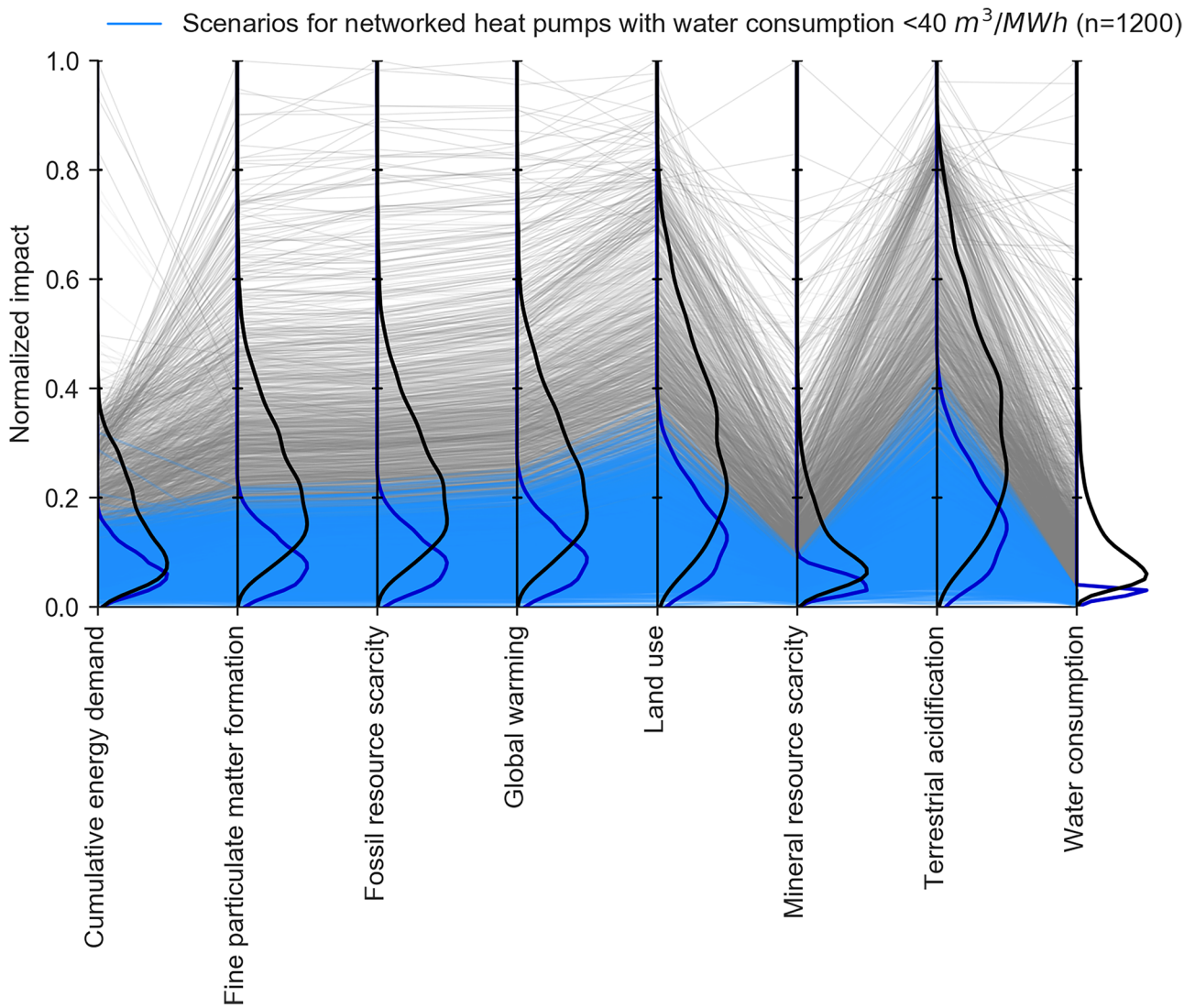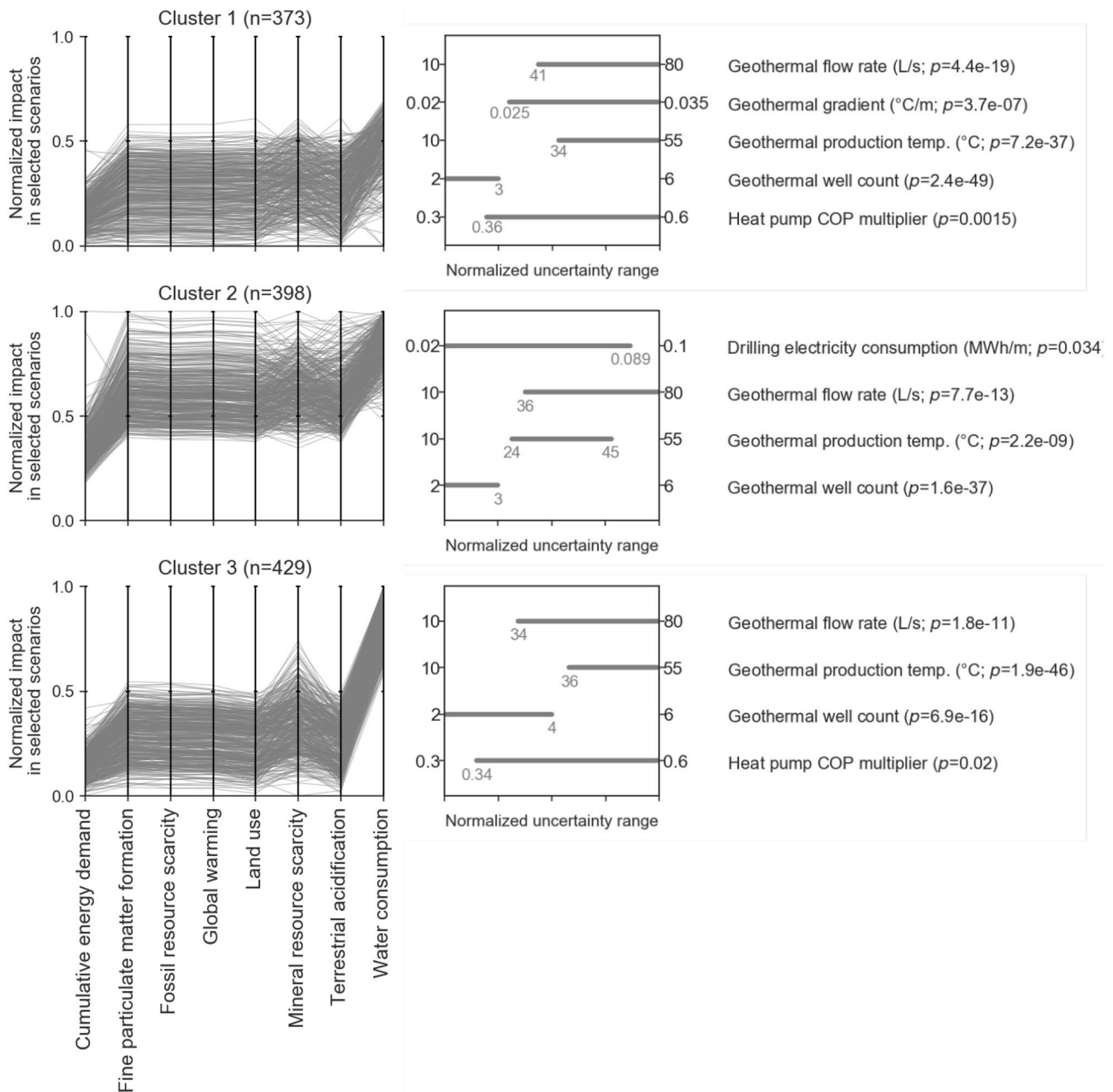
Fig. 6 Subset of scenarios of interest for the networked heat pump alternative (in blue; $n=1200$) and full ensemble of scenarios sampled with Latin Hypercube for this alternative (in gray; $n=12,000$). Impacts are shown using parallel coordinates, with each line representing one scenario. For each impact, outcomes are normalized to [0,1] across the full ensemble. A Gaussian kernel density estimate shows the distribution of outcomes for each impact in the scenarios of interest (in dark blue) and in the full ensemble of scenarios for the networked heat pump alternative (in black)

that these clusters are all associated with a relatively high geothermal flow rate (≥ 37 L/s), combined with a relatively high production temperature (≥ 36 °C). However, opposite portions of the heat pump COP multiplier range are retained in different clusters, leading to different outcomes. The production temperature and COP multiplier did not emerge as highly influential overall for water consumption in the GSA but would nonetheless be decision-relevant in this case. As such, a low range for the COP multiplier (≤ 0.37), combined with a midrange production temperature (36–49 °C), leads to relatively high impacts in cluster 2. Impacts can instead

be minimized as in cluster 1, by combining the upper range of this parameter (≥ 0.4) with a higher production temperature (≥ 39 °C) and a limited number of wells (≤ 3). However, following the definition of the scenarios of interest, the networked heat pump alternative would have a considerably lower water consumption in all these cases, when parameterized with equivalent combinations of the PRIM input ranges that are rescaled to its specific uncertainty bounds. In sum, the networked heat pump alternative may be preferable in situations where a larger geothermal capacity is required and where water consumption is critical.

Fig. 7 Left column: clusters identified using spectral clustering in the subset of scenarios of interest for the networked heat pump alternative. Impacts are shown using parallel coordinates after re-normalizing to [0,1] within the scenarios of interest. Each line represents one scenario. Right column: combinations of uncertainty ranges associated with each cluster, identified using PRIM at a coverage threshold of 75%. The nor-malized uncertainty range shows the full range of input uncertainties sampled in the Latin Hypercube ensemble for the networked heat pump alternative. Gray lines show the restricted range of each input found to be significant ($p < 0.05$) for each cluster. Estimated p values for the significance of each restriction are shown in parenthesis

We conclude the analysis with the direct heating alternative (Online Resource 2, Fig. S21) focusing on scenarios that are Pareto-dominated by the central heat pump alternative across the eight impacts. In relation to the full ensemble of scenarios sampled for this alternative, we find that the scenarios of interest are largely found in the upper tail of the outcome distributions. As with the other alternatives, the clustering step identifies three clusters spanning the range of outcomes in the scenarios of interest (Online Resource 2, Fig. S22). Due to the high correlation between impacts (Online Resource 2, Fig. S4), the relative magnitude of different impacts in each cluster is largely uniform. When only

using the input parameters of the direct heating alternative to explain the clusters, the PRIM algorithm did not reach the coverage threshold of 75% due to the small size of cluster 3 in particular. Given that the scenarios of interest are also defined on the basis of the performance of the central heat pump alternative, we add the latter's heat pump COP parameter to the analysis and find that it emerges as a significant restriction for all clusters. Overall, the Pareto-dominated scenarios are associated with a relatively low flow rate (≤ 42 L/s), a higher number of wells (≥ 4), and a relatively efficient heat pump in the reference central heat pump alternative (COP multiplier ≥ 0.34). This is consistent with the GSA results, which indicated that the first two parameters dominated the uncertainty of most impacts except cumulative energy demand. The worst-case scenarios in cluster 3 are furthermore associated with a low geothermal gradient (≥ 0.025 °C/m), which would require deeper drilling for a given production temperature and greater consequent impacts. As such, in the presence of a low geothermal gradient, combined with a relatively low flow rate, a relatively efficient heat pump, and a large number of wells, the central heat pump alternative would be preferable to direct geothermal heating across all impacts.

5 Discussion and conclusions

This study combined methods for distribution-based GSA, spectral clustering, and scenario discovery to demonstrate the insights which can be gained from these methods in a typical LCA application presenting non-normal outputs, trade-offs across multiple impacts, and interactions between inputs. As such, the PAWN distribution-based sensitivity indices generally identified influential inputs consistently with variance-based first-order and total Sobol indices. However, occasional discrepancies highlighted the potentially misleading interpretation of variance-based indices on non-normal distributions. For instance, the geothermal gradient parameter was found to have a disproportionate impact on the conditional variance of water consumption in the networked heat pump design alternative, whereby fixing the parameter at a certain point of its range could counterintuitively increase the variance of this impact, and thus its uncertainty. This parameter was more prominent using the PAWN indices. Furthermore, differences in the relative values of the median and maximum statistics used to define PAWN indices highlight relevant aspects for further analysis, such as the presence of specific input ranges that may lead to outliers on certain impacts. Nonetheless, the PAWN indices should ideally be used in a "meta-sensitivity analysis" framework (Puy et al. 2020) to test the robustness of these indices to parameters of the method, such as the number of conditioning intervals and sampling type.

In parallel, spectral clustering provides a useful exploratory step in cases where analysts and decision-makers must consider potential trade-offs between multiple impacts. While several impacts were highly correlated in our case study, spectral clustering nonetheless highlighted groups of scenarios presenting trade-offs across the remaining impacts, such as water consumption in the two heat pump alternatives. We then applied the PRIM technique for scenario discovery, identifying specific ranges and combinations of uncertainties associated with these clusters. While second-order Sobol indices pointed toward the overall importance of certain interactions, for instance across geothermal flow rate, production temperature, and number of wells on the water consumption of the networked heat pump alternative, PRIM provided more precise information on the specific values of these parameters that were associated with different patterns of impacts. Conversely, the production temperature and heat pump COP multiplier were found to be decision-relevant in our scenario discovery for the central heat pump alternative, despite having a low overall influence on water consumption using either GSA approach. Scenario discovery thus makes it easier for analysts to precisely identify design parameters or uncertainties associated with specific outcomes across single or multiple impacts, or to compare different design alternatives. Unlike second-order Sobol indices, which require a very large sample size for convergence in typical applications, this approach is useful on smaller ensembles of scenarios. It can also highlight the combinations of multiple input parameters associated with certain output patterns: while Sobol indices are in practice limited to identifying interactions between pairs of parameters, PRIM identified combinations of four or more influential parameters (Fig. 7, right column). While we assumed independent and uniformly distributed inputs in our case study, we note that PRIM can be used regardless of input correlations, or assumptions on input and output distributions. It can similarly be applied on experimental designs that include categorical uncertainties; these can be mapped to structural "switches" in the model to support the combined exploration of parametric and method uncertainties, such as the choice of weighting methods or characterization factors.

We emphasize that we do not present distribution-based GSA or scenario discovery as replacements for variance-based GSA. Except for their dependence on a specific statistical moment, variance-based sensitivity indices otherwise meet all criteria described by Liu and Homma (2009) for "ideal" sensitivity indices, in particular due to their clear mathematical interpretability and their straightforward (if computationally intensive) computation. However, the fundamental assumptions that underpin variance-based GSA should be made explicit in analyses which rely on this approach, by evaluating whether variance is a suitable characterization of

uncertainty. This can be assessed on the basis of the model output distributions by using a statistical test such as Shapiro–Wilk and a quantile–quantile visualization to evaluate the (log-)normality of model outputs. Visualizing the conditional output variance (as in Online Resource 2, Fig. S9) can also illustrate the implications of resolving the uncertainty of a specific parameter: if fixing the parameter at a certain value leads to increased variance, first-order Sobol indices should be interpreted with care. In parallel, the techniques used in this work for distribution-based GSA and scenario discovery can easily be added to variance-based GSA in a multi-method approach, without adding computing costs. As these techniques are applicable with relatively small sample sizes, they are also useful for models for which variance-based GSA is impractical and which may instead benefit from recent advancements in efficient sensitivity estimators (reviewed in Lo Piano et al. 2021). Such a multi-method approach would support a robust assessment of input sensitivities for LCA, while yielding more precise insights into decision-relevant input ranges and interactions between uncertainties. With this paper, we thus hope to highlight the usefulness of emerging methods for the analysis and interpretation of uncertainty in LCA models.

Declarations

Competing interests The authors declare no competing interests.

References

Bisinella V, Conradsen K, Christensen TH, Astrup TF (2016) A global approach for sparse representation of uncertainty in life cycle assessments of waste management systems. Int J Life Cycle Assess 21(3):378–394. https://doi.org/10.1007/s11367-015-1014-4

Borgonovo E (2006) Measuring uncertainty importance: investigation and comparison of alternative approaches. Risk Anal 26(5):1349–1361. https://doi.org/10.1111/j.1539-6924.2006.00806.x

Borgonovo E (2007) A new uncertainty importance measure. Reliab Eng Syst Saf 92(6):771–784. https://doi.org/10.1016/j.ress.2006.04.015

Butler MP, Reed PM, Fisher-Vanden K, Keller K, Wagener T (2014) Identifying parametric controls and dependencies in integrated assessment models using global sensitivity analysis. Environ Model Softw 59:10–29. https://doi.org/10.1016/j.envsoft.2014.05.001

Cucurachi S, Borgonovo E, Heijungs R (2016) A Protocol for the global sensitivity analysis of impact assessment models in life cycle assessment. Risk Anal 36(2):357–377. https://doi.org/10.1111/risa.12443

Dalal S, Han B, Lempert R, Jaycocks A, Hackbarth A (2013) Improving scenario discovery using orthogonal rotations. Environ Model Softw 48:49–64. https://doi.org/10.1016/j.envsoft.2013.05.013

Dorning MA, Diffendorfer JE, Loss SR, Bagstad KJ (2019) Review of indicators for comparing environmental effects across energy sources. Environ Res Lett 14(10):103002. https://doi.org/10.1088/1748-9326/ab402d

Finnveden G (2000) On the limitations of life cycle assessment and environmental systems analysis tools in general. Int J Life Cycle Assess 5(4):229. https://doi.org/10.1007/BF02979365

Friedman JH, Fisher NI (1999) Bump hunting in high-dimensional data. Stat Comput 9(2):123–143. https://doi.org/10.1023/A:1008894516817

Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. Mach Learn 63(1):3–42. https://doi.org/10.1007/s10994-006-6226-1

Groen EA, Bokkers EAM, Heijungs R, de Boer IJM (2017) Methods for global sensitivity analysis in life cycle assessment. Int J Life Cycle Assess 22(7):1125–1137. https://doi.org/10.1007/s11367-016-1217-3

Groves DG, Lempert RJ (2007) A new analytic method for finding policy-relevant scenarios. Glob Environ Change 17(1):73–85. https://doi.org/10.1016/j.gloenvcha.2006.11.006

Hadka D, Herman J, Reed P, Keller K (2015) An open source framework for many-objective robust decision making. Environ Model Softw 74:114–129. https://doi.org/10.1016/j.envsoft.2015.07.014

Heijungs R, Guinée JB, Mendoza Beltrán A, Henriksson PJG, Groen E (2019) Everything is relative and nothing is certain. Toward a theory and practice of comparative probabilistic LCA. Int J Life Cycle Assess 24(9):1573–1579. https://doi.org/10.1007/s11367-019-01666-y

Herman J, Usher W (2017) SALib: An open-source Python library for sensitivity analysis. J Open Source Softw 2(9). https://doi.org/10.21105/joss.00097

Hornberger GM, Spear RC (1981) Approach to the preliminary analysis of environmental systems. J Env Mgmt 12(1):7–18

Huijbregts M (1998) Application of uncertainty and variability in LCA. Int J Life Cycle Assess 3(5):273. https://doi.org/10.1007/BF02979835

Huijbregts M, Steinmann Z, Elshout P, Stam G, Verones F, Vieira M, Hollander A, Zijp M (2017) ReCiPe 2016 v1. 1-A harmonized life cycle impact assessment method at midpoint and endpoint level: Report I. Characterization (No. RIVM Report 2016–0104a). Natl Inst Public Health Environ Bilthoven Neth

International Organization for Standardization (2006) Environmental management: life cycle assessment; requirements and guidelines. ISO Geneva

Itten R, Frischknecht R, Stucki M (2014) Life cycle inventories of electricity mixes and grid. treeze Ltd., Uster, Switzerland

Jaxa-Rozen M, Bloemendal M, Rostampour V, Kwakkel J (2016) Assessing the sustainable application of aquifer thermal energy storage. In: European Geothermal Congress. Strasbourg, France

Jaxa-Rozen M, Kwakkel J (2018) Tree-based ensemble methods for sensitivity analysis of environmental models: a performance comparison with Sobol and Morris techniques. Environ Model Softw 107:245–266

Jaxa-Rozen M, Pratiwi AS, Trutnevyte E (2021) Analysis workflow for sensitivity analysis and scenario discovery. https://doi.org/10.5281/zenodo.4201064

Kwakkel JH (2017) The Exploratory Modeling Workbench: an open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. Environ Model Softw 96:239–250

Kwakkel JH, Haasnoot M (2019) Supporting DMDU: a taxonomy of approaches and tools. In: Marchau VAWJ, Walker WE, Bloemen PJTM, Popper SW (eds) Decision Making under Deep Uncertainty: From Theory to Practice. Springer International Publishing, Cham, pp 355–374

Kwakkel JH, Jaxa-Rozen M (2016) Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes. Environ Model Softw 79:311–321. https://doi.org/10.1016/j.envsoft.2015.11.020

Lacirignola M, Blanc P, Girard R, Pérez-López P, Blanc I (2017) LCA of emerging technologies: addressing high uncertainty on inputs' variability when performing global sensitivity analysis. Sci Total Environ 578:268–280. https://doi.org/10.1016/j.scitotenv.2016.10.066

Lempert R, Bryant B, Bankes S (2008) Comparing algorithms for scenario discovery. RAND Corporation, Santa Monica, CA

Liu H, Chen W, Sudjianto A (2005) Relative entropy based method for probabilistic sensitivity analysis in engineering design. J Mech Des 128(2):326–336. https://doi.org/10.1115/1.2159025

Liu Q, Homma T (2009) A new computational method of a moment-independent uncertainty importance measure. Reliab Eng Syst Saf 94(7):1205–1211. https://doi.org/10.1016/j.ress.2008.10.005

Lo Piano S, Ferretti F, Puy A, Albrecht D, Saltelli A (2021) Variance-based sensitivity analysis: the quest for better estimators and designs between explorativity and economy. Reliab Eng Syst Saf 206:107300. https://doi.org/10.1016/j.ress.2020.107300

Mahbub N, Oyedun AO, Zhang H, Kumar A, Poganietz W-R (2019) A life cycle sustainability assessment (LCSA) of oxymethylene ether as a diesel additive produced from forest biomass. Int J Life Cycle Assess 24(5):881–899. https://doi.org/10.1007/s11367-018-1529-6

Marchau VAWJ, Walker WE, Bloemen PJTM, Popper SW (eds) (2019) Decision making under deep uncertainty: from theory to practice. Springer International Publishing

Michiels F, Geeraerd A (2020) How to decide and visualize whether uncertainty or variability is dominating in life cycle assessment results: a systematic review. Environ Model Softw 133:104841. https://doi.org/10.1016/j.envsoft.2020.104841

Moallemi EA, de Haan F, Kwakkel J, Aye L (2017) Narrative-informed exploratory analysis of energy transition pathways: a case study of India's electricity sector. Energy Policy 110:271–287. https://doi.org/10.1016/j.enpol.2017.08.019

Muller S, Mutel C, Lesage P, Samson R (2018) Effects of distribution choice on the modeling of life cycle inventory uncertainty: an assessment on the Ecoinvent v2.2 Database. J Ind Ecol 22(2):300–313. https://doi.org/10.1111/jiec.12574

Noacco V, Sarrazin F, Pianosi F, Wagener T (2019) Matlab/R workflows to assess critical choices in global sensitivity analysis using the SAFE toolbox. MethodsX 6:2258–2280

Nossent J, Elsen P, Bauwens W (2011) Sobol sensitivity analysis of a complex environmental model. Environ Model Softw 26(12):1515–1525. https://doi.org/10.1016/j.envsoft.2011.08.010

Payen S, Basset-Mens C, Perret S (2015) LCA of local and imported tomato: an energy and water trade-off. J Clean Prod 87:139–148. https://doi.org/10.1016/j.jclepro.2014.10.007

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

Pianosi F, Beven K, Freer J, Hall JW, Rougier J, Stephenson DB, Wagener T (2016) Sensitivity analysis of environmental models: a systematic review with practical workflow. Environ Model Softw 79:214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

Pianosi F, Wagener T (2015) A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. Environ Model Softw 67:1–11. https://doi.org/10.1016/j.envsoft.2015.01.004

Pianosi F, Wagener T (2018) Distribution-based sensitivity analysis from a generic input-output sample. Environ Model Softw 108:197–207. https://doi.org/10.1016/j.envsoft.2018.07.019

Prado-Lopez V, Wender BA, Seager TP, Laurin L, Chester M, Arslan E (2016) Tradeoff Evaluation improves comparative life cycle assessment: a photovoltaic case study. J Ind Ecol 20(4):710–718. https://doi.org/10.1111/jiec.12292

Pratiwi AS, Trutnevyte E (2021) Life cycle assessment of shallow to medium-depth geothermal heating and cooling networks in the State of Geneva. Geothermics 90:101988. https://doi.org/10.1016/j.geothermics.2020.101988

Pujol G, Iooss B, Janon A (2017) Sensitivity: global sensitivity analysis of model outputs.

Puy A, Lo Piano S, Saltelli A (2020) A sensitivity analysis of the PAWN sensitivity index. Environ Model Softw 127:104679. https://doi.org/10.1016/j.envsoft.2020.104679

Qin Y, Suh S (2017) What distribution function do life cycle inventories follow? Int J Life Cycle Assess 22(7):1138–1145. https://doi.org/10.1007/s11367-016-1224-4

Ravalico JK, Maier HR, Dandy GC (2009) Sensitivity analysis for decision-making using the MORE method—a Pareto approach. Reliab Eng Syst Saf 94(7):1229–1237. https://doi.org/10.1016/j.ress.2009.01.009

Razali NM, Wah YB (2011) Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. J Stat Model Anal 2(1):21–33

Reap J, Roman F, Duncan S, Bras B (2008) A survey of unresolved problems in life cycle assessment. Int J Life Cycle Assess 13(5):374. https://doi.org/10.1007/s11367-008-0009-9

Rosenbaum RK, Georgiadis S, Fantke P (2018) Uncertainty Management and Sensitivity Analysis. In: Hauschild MZ, Rosenbaum RK, Olsen SI (eds) Life cycle assessment: theory and practice. Springer International Publishing, Cham, pp 271–321

Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Saltelli A (2002) Sensitivity analysis for importance assessment. Risk Anal 22(3):579–590. https://doi.org/10.1111/0272-4332.00040

Saltelli A, Annoni P (2010) How to avoid a perfunctory sensitivity analysis. Environ Model Softw 25(12):1508–1517. https://doi.org/10.1016/j.envsoft.2010.04.012

Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S (2010) Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput Phys Commun 181(2):259–270. https://doi.org/10.1016/j.cpc.2009.09.018

Saltelli A, Tarantola S, Campolongo F, Ratto M (2004) Sensitivity analysis in practice: a guide to assessing scientific models. Wiley Online Library

Savage LJ (1951) The theory of statistical decision. J Am Stat Assoc 46(253):55–67

Sobol IM (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math Comput Simul 55(1–3):271–280. https://doi.org/10.1016/S0378-4754(00)00270-6

Spear RC, Grieb TM, Shang N (1994) Parameter uncertainty and interaction in complex environmental models. Water Resour Res 30(11):3159–3169. https://doi.org/10.1029/94WR01732

Steinmann P, Auping WL, Kwakkel JH (2020) Behavior-based scenario discovery using time series clustering. Technol Forecast Soc Change 156:120052. https://doi.org/10.1016/j.techfore.2020.120052

Vidal R, Sánchez-Pantoja N (2019) Method based on life cycle assessment and TOPSIS to integrate environmental award criteria into green public procurement. Sustain Cities Soc 44:465–474. https://doi.org/10.1016/j.scs.2018.10.011

von Luxburg U (2007) A tutorial on spectral clustering. Stat Comput 17(4):395–416

Wei W, Larrey-Lassalle P, Faure T, Dumoulin N, Roux P, Mathias J-D (2015) How to Conduct a Proper Sensitivity Analysis in Life Cycle Assessment: taking into account correlations within LCI data and interactions within the LCA calculation model. Environ Sci Technol 49(1):377–385. https://doi.org/10.1021/es502128k

Wernet G, Bauer C, Steubing B, Reinhard J, Moreno-Ruiz E, Weidema B (2016) The ecoinvent database version 3 (part I): overview and methodology. Int J Life Cycle Assess 21(9):1218–1230. https://doi.org/10.1007/s11367-016-1087-8

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.