# Introduction

Computational modelling is a relatively young field which was constrained in its early years by the disproportionately high cost and limited availability of computational power. Since then computers have become ubiquitous, with vast amounts of available computational power and speed. Perhaps most importantly from a financial perspective, computational cost has now become cheaper than that of highly trained researchers. This has led computational modelling to become a key activity in many areas of research, whilst enabling an ever-increasing level of complexity to be modelled.

Here the level of complexity is reflected by the number of parameter factors a model takes as inputs. Parameter factors may also be referred to as ‘parameters’, ‘factors’ or simply ‘inputs’ in the literature (Norton, 2015). Increasing the number of parameters allows for a more detailed representation of the system being investigated, whilst also increasing computational cost and model complexity, often at an exponential rate. This is referred to as the curse of dimensionality, where each parameter represents a dimension. Increased detail (and thus complexity) may not always be justified or necessary with respect to the aims of the modeling exercise. Such complex models are described as being ‘over-parameterized’ (van Griensven et al., 2006).

Uncertainty and Sensitivity Analysis (UA/SA) helps modelers identify the relative importance of each parameter factor which influences model results within a given problem context. Each parameter factor may influence one or more outputs and could (conditionally) affect the importance of other factors; referred to as parameter interaction. Factors may be ‘insensitive’ – having little to no effect on model results – and can then be fixed (made static and unchanging) or removed to reduce complexity. Understanding the relative importance of factors that are ‘sensitive’ can aid in the development of better monitoring strategies and experiment design by, as an example, indicating the priority of data to be collected and how much is necessary. Limiting complexity helps to constrain the parameter space, which has many benefits. It can aid in reducing model runtime by eliminating the need to consider insensitive parameters, which then helps ease the computational cost of optimizing the model, and is a key activity in developing surrogate models.

The influence of uncertain parameters and how this impacts decision-making can be surveyed through Exploratory Modelling and Analysis (EMA). EMA could be described as the exploration of conditions represented through scenarios and conditions. These scenarios consider parameter and structural uncertainties in order to generate predictions under future conditions. SA and sampling techniques (such as MC, factorial methods, and optimization techniques) can be used in conjunction with EMA to great effect, especially when only partial information about a system is available (Kwakkel and Pruyt, 2013).

UA/SA may also be used as part of a quality assurance step. When applied in such a manner UA/SA can indicate, through statistical measures, questionable outputs due to data quality issues or bugs in the modelling software. Unexpected outputs may also be a result of parameter interactions and combinations that may not be possible. It is not suggested here that UA/SA is a complete substitute for code/model testing – rather it is a complement that further reinforces the validity of the model.

The importance of understanding how the model represents the problem frame is often underappreciated for a variety of reasons. Consequently, UA/SA processes may not be applied to an appropriate degree (Jakeman et al., 2006; Norton, 2015; Saltelli and Annoni, 2010)**.** Modelers may attempt to optimize and calibrate models without the requisite knowledge and understanding of the model behavior. Donald Knuth, a well-known computer scientist, once stated that premature optimization is the root of all evil (Knuth, 1974). This was in reference to software developers who often spend a lot of time and energy on modifying code to achieve better computational performance, sometimes without a clear understanding of the implications. We find that this is equally true in the context of modelling – one should not optimize a model without an adequate understanding of the role parameter factors play. UA/SA processes can help in this regard.

A significant volume of work has been conducted in the field of UA/SA. An initial query on the Web of Science database resulted in over 500,000 publications across all fields – far too many to comprehensively review, at least in a timely manner. In this paper we aim to provide 1) an overview of UA/SA research trends within the field of environmental modeling, and 2) an introductory guide to the available software tools and packages for those new to the field of environmental modelling and analysis. These goals are achieved through a hybrid bibliometric approach.

# The (hybrid) bibliometric approach

The collection of publications (the ‘corpora’) were gathered from Clarivate Analytics’ Web of Science (WoS) database using the WoS web-based Application Programming Interface (API). Use of the API enabled access to the publication data and metadata including titles, abstract text, author supplied keywords, and DOIs. The overarching subject areas and keywords included in the search parameters are specified in **Error! Reference source not found.**. Publications in the resulting corpora were taken to represent the field of uncertainty and sensitivity analysis in the field of environmental modeling. The general search and analysis approach is depicted in **Error! Reference source not found.**.

The collection of publications was iteratively and incrementally refined through a semi-autonomous process. Topic modelling was applied to cluster the publications based on semantic similarity, which aided in identifying relevant texts. Analysis of the trends and citations within relevant topics and their representative texts was conducted to construct an overview of the general research trend and identify software tools and packages. Additional complementary publications were selected based on the author’s own judgement to further the discussion. Further discussion of the approach is given in the following sub-sections.

## Initial search

The initial corpora for the analysis was identified by specifying the search phrase **[TS=]**. Only English language publications were considered for this study. Further details of the search phrase and the specific keywords used are given in **Error! Reference source not found.** and **Error! Reference source not found.**. A time frame limit of 2000 to 2017 was put in place as the initial search resulted in **[over 500,000]** matches. Because scientific research is largely additive in nature (i.e. we build on prior works) and there is an exponential growth of published material (Bornmann and Mutz, 2015; Haddaway and Westgate, 2018), we assume in this analysis that most, if not all, relevant information is represented by the publications in the corpora. Data was retrieved with the use of Wosis, a python package developed in-house to simplify the process of querying the WoS database and aid in data analysis and visualization (**[ref package DOI]**).

The initial corpora from WoS consisted of **[13 692]** publications from which documents without a valid DOI (**429 publications**) were removed. Further to this, publications which included “Proceedings” in the journal name were removed, and those corresponding to the keywords “life cycle assessment,” “product life cycle,” and “image processing.” Unrelated journals were excluded, and journals with less than three publications in the corpora. The final corpora consisted of **[11718]** publications.

The publications in the field have been increasing at an exponential rate (depicted in **Error! Reference source not found.**) with Journal of Hydrology having the most publications overall and experiencing the largest year-on-year gain within the analyzed time frame. The 10 most cited papers from across the top 5 journals (by number of publications) were then identified and collated (see **Error! Reference source not found.**). Most were published in either the Journal of Hydrology (four in total) or Environmental Modelling and Software (five publications), with the single exception published in Structural and Multidisciplinary Optimization.

## Topic Identification

Topic models attempt to cluster the corpora into similar or related topics. They group texts by the words occurring commonly in each text, relying on the assumption of common semantics in various fields. Topic models can be more efficient at discovering texts relevant to a field compared to an exhaustive manual search as it identifies relevant texts based on commonly occurring words which indicate a specific context or topic. For example, sensitivity in the context of sensitivity analysis would be expected to appear in texts containing words such as “analysis”, “uncertainty”, and “modelling”, whereas it could also appear in relation to perceptiveness of technologies to stimulus, in which case it would appear with such words as “precision,” “technology,” “response.” Through this approach a semi-supervised search can be conducted wherein the topic model is applied, irrelevant topics filtered, and the process repeated until only the relevant texts are left.

Topic modeling can be of great benefit to scientific literature reviews. (Achakulvisut et al., 2016) outline the shortcomings of a standard keyword search for scientific literature: a likelihood of bias, Matthew’s effects, the specialized nature of scientific fields, the need for automation as the literature grows increasingly, and the rapid emergence of new fields. It is becoming increasingly difficult for researchers to communicate and apply the available knowledge, given the sheer volume of papers and corresponding hours needed for manual sorting (Westgate et al., 2018). Topic modelling, in taking advantage of machine learning, can greatly reduce the required hours for a thorough, less-biased, and more systematic review of a subject area. A key advantage, for example, is that it can analyze the data stored in abstracts, a richer source of information than keywords.

In this study a Non-negative Matrix Factorization (NMF) was applied to model the topics within the corpora. NMF is a linear-algebraic algorithm which requires the number of topics to be pre-specified by the modeler. A publication may be assigned to one or more topics with the NMF approach. The NMF model requires a document-term matrix, also referred to as a bag-of-words matrix, in which each row represents a document, and each column vector represents a considered word token. A complete explanation of the NMF approach is not in the scope of this paper, however further information can be found in the following references (Arora et al., 2012).

For our purposes the word tokens consisted of the text found in the title, abstract, and keywords. Term Frequency-Inverse Document Frequency ranking (TF-IDF) was used to select 1000 tokens from these texts. A high TF-IDF score indicates that the word has a high frequency within specific document(s), but a low number of occurrences within the entire corpora. Weighting the score in such a manner has the effect of filtering out commonly used tokens which may not have high semantic importance (such as “the”, “and”, etc). TF-IDF is a common ranking method used in text mining (Beel et al., 2016).

Topic modeling and bibliometric analysis have been applied before to reduce the time and difficulties encountered when conducting systematic reviews (Westgate and Lindenmayer, 2017). Although software is available to aid in these bibliometric approaches, currently no single software package provides all necessary functionality. Arguably the application of conjunctive systematic mapping and bibliometric analysis is still in its infancy (as evidenced by Nakagawa et al., 2018).

### Citation Analysis

Citation analysis indicates the papers being referred to by other papers within the corpora as well as the overall number of citations the given publication has received, the assumption here being that impactful papers are more likely to be cited. The number of citations is then used to indicate papers that are of high importance to the subject at hand.

### Trend Analysis

Analyzing the trend of publications within topic areas can aid in identifying the general focus and direction taken by the research community.

### Key Phrase Identification

The corpora can be further constrained through key phrase identification, which the wosis package can apply. The automated process identifies and displays a pre-selected number of key phrases for each given publication aiding reviewers to identify irrelevant publications. The approach implemented in wosis attempts to identify phrases of interest by scoring sentences based on its similarity with other sentences throughout the abstract text.

Sentences with three or less tokens (i.e. words, numbers, or other; counted by splitting the text on individual spaces) are ignored. Sentence similarity is scored based on the ratio of the intersection of two sentences. This functionality is provided by the Python `fuzzywuzzy` string matching package through the `token\_set\_ratio()` function. Candidate sentences are initially filtered based on the presence of a root word – taken to be the token that appears in the middle of a candidate phrase; an approach used in Rabby et al. (2018).

This approach assumes that important features of the publication (such as its findings) will be introduced or framed and subsequently discussed in the abstract text. The implemented approach is therefore dependent on the abstract length, with longer texts preferred and poor performance can be expected for very short abstracts (e.g. no key phrases could be identified for abstracts less than three sentences long). Comparisons with RAKE – Rapid Automatic Keyword Extraction (Rose et al. 2010) – implemented through the rake-nltk python package indicates that the above approach produces, subjectively, key phrases that are more relevant for the purpose of this study.

[**TODO**: Table or figure comparing the results from RAKE algorithm vs the custom approach implemented in wosis]

# Key terminologies and methods

For local SA (LSA) the partial derivatives of output with respect to each input parameter are computed at one point in the sample space to determine sensitivity indices. The simplicity of the procedure is advantageous, as well as being computationally inexpensive for first order derivatives. However, the method only provides a robust SA for linear or additive models (Saltelli and Annoni, 2010): it doesn’t account for parameter interactions and becomes computationally expensive when higher order effects are considered.

Global sensitivity analysis methods can be classed into two broad categories – derivative- and variance-based. Derivative-based GSA methods provide indices which characterise the distributional properties of partial derivatives (Razavi et al., 2019). Partial derivatives are computed for a selection of sample points (as opposed to one base point) and are combined.

Variance-based approaches determine how different factors contribute to model variance by analysing and decomposing the variance in model outputs (Razavi et al., 2019).

The strength of GSA methods comes in that they provide a more robust depiction of model uncertainty, in particular by comprehensively accounting for parameter interactions.

A weakness is computational cost and an assumption of random distribution of output values in the parameter space.

A third category of methods was proposed by (Razavi et al., 2019). These variogram-based methods characterise sensitivity by generating directional variograms associated with each parameter. The methods account for the spatial distribution of the model response.

When implementing a GSA, sampling methods are required to adequately represent parameter space whilst limiting computational expense. OAT and LSA do not require sampling. For GSA, MC sampling is commonly used, although it offers a limited representation of the total space. LHS resolves this issue (Norton, 2015). A weakness of the LHS method is that the single sample it generates cannot be enlarged without altering the distributional properties: the user should be aware of the required sample size before implementing the method (Razavi et al., 2019). LHS is an example of a stratified sampling method, and these methods, as well as MC and Sobol’ sequences, are used for variance-based GSA. Other commonly used sampling techniques are those designed specifically for particular SA methods, such as FAST, Morris OAT, and Sobol’ sensitivity indices (Norton, 2015).

# Typical sensitivity analysis process

SA methods determine where the uncertainty in a model is coming from (Saltelli and Annoni, 2010). Typically, the analysis identifies factors or groups of factors that have the greatest impact on model uncertainty. Regardless of the analysis being conducted, whether it be SA, UA, or EM, the primary conceptual idea is that of mapping a selection of model inputs to their results (i.e. associate the inputs to their resultant outputs). **[explanation of ua/sa/em difference]**

(Pianosi et al., 2015) identify three stages in a sensitivity analysis: selecting a sample of input values from the variability space, running a model evaluation against these input values, and applying a sensitivity analysis method to the input/output samples to compute sensitivity indices.

Here, the *variability space* refers to all possible combinations of values that can be assigned to a model’s input parameter set. By running the model with the values sampled from the variability space and taking note of the resultant outputs, analyses can be conducted to calculate the influence that a specific input, or set of inputs, may have; their sensitivities. The sample of inputs, ideally, would be suitably inclusive as to capture the entire range of possible parameter values without excessive computational cost. Knowledge of the probability density function of the model parameters is also useful, to focus the analysis on the regions with the highest probability (Norton, 2015). Random sampling with a uniform probability density over the parameter range is commonly assumed, but this assumption lacks accuracy (see Norton, 2015).

Common sampling techniques include Monte Carlo sampling, Latin hypercube sampling, and the Morris screening method. After evaluating the model over the sampled inputs, a sensitivity analysis is conducted. A local sensitivity analysis identifies the importance of a factor’s influence on model output by examining the partial derivative of the output with respect to the factor. The term local refers to these derivatives being taken with respect to a single point in parameter space. Such a SA method will produce a robust inference for linear or additive models (Saltelli and Annoni, 2010). OAT SA involves perturbing one model factor, and examining the resulting output, whilst fixing all other factors. More generally, a global sensitivity analysis examines the variation in output resulting from variation across all the inputs.

# Overview of recent developments

Recent trends in sensitivity analysis suggest a shift from a previous deference to local sensitivity methods. Prior to 2010, ‘one-factor-at-a-time’ (OAT) local SA was the most prevalent practice in the literature (Saltelli and Annoni, 2010). Our analysis showed an increased interest in GSA methods, occurring after 2010 (**Fig.**). It seems modelers have come to realize the greater importance of higher quality sensitivity analyses, especially for improving the reliability of forecasting, prediction, and policy making. (Saltelli and Annoni, 2010) demonstrated the inefficacy of OAT analyses using a geometric proof, wherein it was shown that simply varying the parameter values one at a time cannot adequately capture the output space. They argued that such OAT analyses are, in all but the simplest (linear/additive) cases, unsuitable. Whilst it is not possible to ascertain the direct impact of this paper on current sensitivity analysis practices, it at least reflected the changing attitudes towards SA which were developing in the modelling community.

In the past decade, there has been an increasing awareness of the shortcomings of local SA, and more efficient, comprehensive UA/SA techniques and approaches have been sought. Developers search for the best compromise between global sensitivity measures and computational efficiency. There has been a recently increased interest in global SA based on design of experiments, as these methods are recognized as the currently best available compromise (Gan et al., 2014). The bibliometric analysis shows an increased use of global SA methods, with this shift in interest occurring in the period 2005-2010. Local SA and OAT methods are still in use, however. It is difficult to ascertain the extent of continued OAT analysis in the modelling community through keyword analysis, as most researchers applying this technique will not explicitly refer to it as OAT. A keyword-match was run over the filtered corpora, using the keywords “local sensitivity”, “OAT”, “one-at-a-time”, and “perturb”. The resulting papers were **[751]** in number. As a percentage of the corpora, these papers are decreasing.

Having such a volume of publications returned by the search, a keyword-match was used to sort them into their respective fields. Fields searched for were hydrology, chemical engineering, mechanical engineering, ecology, risk assessment, technology, agriculture, computational physics, computer science, statistics, and mathematical modeling. The fields were chosen based on the top 20 journals by publication. All fields, except hydrology, showed a decreasing trend as a percentage of the corpora. Hydrology had an increase of roughly 1% over the timeframe. The top 5 fields by publication were chemical engineering (146), hydrology (137), technology (121), ecology (102), and agriculture (78). One must bear in mind the limitations to such analysis, as some returned papers may have been those condemning OAT: the top cited paper from the top journals by publication was (Saltelli and Annoni, 2010).

A criteria-matching search for best practices revealed increased interest in the years 2014-2016. Only 25 papers returned the search query. A key-phrase extraction algorithm facilitated identifying the fields publishing in this area. The largest number of publications concerned model validation and development (11, especially regarding environmental contexts: hydrology and chemistry), followed by environmental risk assessment (8, including emissions), SA (3), life cycle assessment (2, regarding sustainable production and consumption), and legislation (1, regarding climate change and ecology).

Three SA packages, released within the past five years, reflect these changing attitudes: PSUADE (Gan et al., 2014), SAFE (Pianosi et al., 2015), and VARS-TOOL (Razavi et al., 2019).

PSUADE (a Problem Solving Environment for Uncertainty Analysis and Design Exploration) provides users with implementations of uncertainty quantification methods, including sampling techniques and SA methods (both local and global). The package has had general application to various modelling scenarios.

SAFE (Sensitivity Analysis For Everybody) provides users with implementations of global SA methods, with the ability to perform multiple SAs, robustness assessment, and convergence analysis without further model runs. As reflected in its name, this package was designed to allow global SA to be accessible to a more general audience.

VARS-TOOL provides implementations of sampling techniques and global SA methods, including derivative-, variance-, and variogram-based, which can all be performed from a single sample. The developers claim that the variogram-based SA approach links both local and global approaches to SA. Variogram-based methods determine which parameters have the most controlling influence on the model output, and hence the model uncertainty.

Furthermore, alternative methods for handling uncertainty have been developed, especially to handle scenarios in which there is large uncertainty, but in which accurate predictions are necessary for future policy making. In such cases, one proposed approach is EMA, which uses computational experiments to systematically explore the consequential outcomes of model uncertainty (Kwakkel and Pruyt, 2013). Rather than simply minimizing uncertainty, uncertainty is treated as inevitable, and responses can be planned to respond to the various feasible outcomes that the model could produce.

## Identified Topic(al) areas

Running a topic model on the filtered corpora enabled us to refine the corpora to those publications most relevant to our analysis, the idea being that the topic model would be run until all the produced topics were relevant.

A topic model was run on the filtered corpora, obtaining 5 topic areas, two of which were of interest to this study. Topics 1 to 3 were considered relevant to SA, optimization, and UA, respectively. The topics had the keywords “model parameters sensitivity models data analysis parameter flow calibration time

,” “optimization design method shape topology problem sensitivity element structural finite

,” and “uncertainty stochastic quantification carlo monte bayesian method uncertainties random polynomial

,” respectively. The three topics were combined and a further topic model was run on the combined collection. This produced five topics, of which Topic 1 appeared relevant to SA, Topic 2 to optimization, and Topics 3 and 4 to UA. The topics had the keywords “model sensitivity parameters analysis data flow models parameter soil results,” “optimization design shape topology method sensitivity structural problem analysis element,” “uncertainty bayesian uncertainties carlo monte model quantification models analysis data,” and “stochastic random method polynomial chaos quantification expansion equations collocation solution,” respectively.

Combining these topics and running a topic model produced 5 topics, again Topic 1 appeared relevant to SA, Topic 2 to optimization, and Topics 3 and 4 to UA. The keywords were now “model sensitivity parameters analysis data models parameter calibration based soil,” “optimization design shape topology method sensitivity structural problem analysis element,” “uncertainty bayesian uncertainties carlo monte quantification analysis models assessment probabilistic,” and “stochastic random polynomial method chaos quantification expansion equations problems collocation,” respectively. A further topic model was run on this combination of topics, but the resulting topics all seemed relevant to the aims of this paper.

These topics were combined (the resulting collection will be referred to as “combined topics”), and a keyword-match was run to sort the publications into those relevant to UA, SA, and optimization. The resulting collections contained 1940, 2751, and 1360 publications, respectively.

Topic modelling the filtered corpora helped to create a collection of papers more relevant to the study (UA/SA). The chosen number of topic models and topics seemed optimal for filtering the given corpora whilst not restricting too far.

A keyword-matching algorithm was applied to the combined topics, identifying those papers relevant to frameworks and guidelines for UA/SA, and those relevant to applications of UA/SA. This produced two collections, which will be referred to as “Frameworks” and “Applications,” respectively. After the first citation analysis of these papers, it was clear that there was significant overlap between the categories, and in particular, the papers that overlapped were more relevant to Frameworks than Applications. Hence, after the first keyword match, the papers matching to Frameworks were removed from the combined subtopics before performing a keyword-match for applications on the remaining papers in the combined collection.

Papers relevant to software implementation were also of interest to this paper. In order to build a corpus of software-related publications, a keyword-matching algorithm was first run on the filtered corpora. The chosen keywords were “software” and “toolkit.” These keywords, through preliminary searches, were found to be most relevant: other terms, such as package, were found to be superfluous or to pick up too many irrelevant publications. This yielded **[~371]** papers. A key-phrase extraction algorithm was run on this collection, which more easily facilitated manual sorting. Manual sorting brought the corpus down to **[201]** papers. Papers were chosen as relevant if they included direct reference to UA/SA or optimization software packages; were theory, review, or framework papers which recommended software implementation to a given field; or referred to other methods and packages of interest to expert opinion, those being GLUE, PEST, PEST++, MODFLOW, HYDROMAD, and SWAT.

Many of the omitted papers included applications of perturbation SA (ie OAT) to what were usually modelling and simulation software packages.

## Research Trends and Directions

A trend analysis of the UA, SA, and optimization topics shows a yearly increase for all three topics, however SA and optimization saw a yearly decrease as a percentage of the combined topics.

The Applications and Frameworks collections both show yearly increases, however the Applications papers decrease proportional to the combined topics.

For Frameworks papers, the top 5 journals by publication all showed an increasing publication trend. These journals were Structural and Multidisciplinary Optimization, Journal of Computational Physics, Environmental Modelling & Software, and Journal of Hydrology.

For Application papers, the journal publication trend of the top 5 journals by number of publications shows an increase in publications for those papers coming from Computer Methods in Applied Mechanics and Engineering, Structural and Multidisciplinary Optimization, Journal of Computational Physics, and Journal of Hydrology. Publications from the International Journal for Numerical Methods in Engineering showed only a slight increase over the timeframe.

Further to this, the 10 most cited papers from across the top 5 journals for both topics were identified and collated. For Frameworks, the 10 most cited papers came from Environmental Modelling & Software (3), Structural and Multidisciplinary Optimization (3), Journal of Hydrology (2), Journal of Computational Physics (1), and Computer Methods in Applied Mechanics and Engineering (1). For Applications, the 10 most cited papers came from Journal of Hydrology (5), International Journal for Numerical Methods in Engineering Hydrological Processes (2), and Journal of Computational Physics (3).

Co-author citation networks were produced using the networkx package, in order to indicate which fields are collaborating in each of the topic areas. **[need to find a way to save the graphs to a file, only option in documentation is .csv, could screenshot]**

**(Trend phases)**

# Sensitivity analysis packages

In order to create a collection of papers pertaining to software packages, the first step was to perform a criteria match on the corpora with the terms “software” and “packages.” This reduced the total corpora to **[468]** publications. Next, a key-phrase extraction algorithm was applied to these publications, and these were manually sorted (made easier with the key-phrase extraction) into a final collection of relevant papers, consisting of **[201]** publications.

The use of software packages in UA/SA and optimization has seen a stable trend over the surveyed timeframe. Proportional to the corpora, there was a peak in publications in this topic area in 2007.

A criteria-matching algorithm was applied to this collection of packages in order to allocate papers to relevant categories, these being Policy, Agriculture, Environment, Water, Socio-economic, Medicine, and Chemistry. Using the key-phrase algorithm, we then searched to see which packages were being used by which fields, and for what purpose.

Papers relating to Policy included risk assessment, hydrological, environmental hazard, and waste management, population viability, and life cycle assessment. Papers included also addressed issues of modelling with insufficient data or data with deep uncertainty. They were concerned with UA/SA, calibration, and decision analysis, especially regarding complex models.

Papers in the Agriculture category addressed emissions, irrigation, biogeochemical models, and agriculture-related life cycle assessments. The papers were concerneed with SA, model development, and calibration.

Papers that met the Water criteria covered a variety of topics: drainage, water quality, groundwater, catchment management, storm discharge, stream-gauging data, infiltration, and flood forecasting. The packages mentioned were SWAT, DUET-H/WQ, PEST, SAKE, TOPMODEL, hydroPSO, SIMLAB, WaterRAT, GANetXL, MODEFLOW, UTCHEM, MULINO-DSS, ArCNLET, and PERSiST. The papers were concerned with model design, development, and implementation, decision analysis and long-term planning, system integration, data recording, assimilation, and generation, UA/SA, uncertainty quantification (UQ), optimization, bootstrap confidence intervals, educational models, and probabilistic projection.

Papers in the Environment category concerned ice-sheet mass, emissions, aerodynamics, soil, life cycle assessments, technologies, and population viability. The packages included were Day Cent, Ecological Risk-O-Meter, and DUE. This category was concerned with UA/SA, UQ, calibration, model design and integration, risk assessment, optimization, and assessing and simulating uncertain variables.

Papers that met the Socio-economic criteria concerned bio-oil, hydropower installation, urban drainage, renewable energy systems and sources, managed hospital care, life cycle assessment, and setup cost. These papers were concerned with techno-economic analysis, SA, model design and parallelization, optimization, feasibility analysis, and enviro-economic analysis.

Papers that met the Medicine criteria concerned heavy metal concentration in drinking water, cardiovascular analysis, cancer risk, and noise impact. These papers focused on spatial analysis, (health) risk assessment, frameworks, modelling, and UA.

Finally, those meeting the Chemistry criteria included biogeochemical models, tracer concentration, soil solution concentration, chemical kinetic systems, chemical persistence and long-range transport, (sustainable) chemical processes, bio-oil upgrading, plasma-assisted methane steam reforming, and wastewater. The packages mentioned were Day Cent, ChemKIN-PRO, TAMkin, and ArcNLET. They concerned modelling, calibration, UA/SA, parameter screening, decision analysis, optimization, and retrofit analysis.

Some first thoughts:

Sensitivity analysis (SA) packages released in the last ten years appear to be more concerned with best practices in modelling and software design. The purposes of these packages are more generalized, and the packages themselves offer a more comprehensive toolkit for uncertainty quantification (UQ). (Pianosi, et al, 2015) outline three principles of good practice for a sensitivity analysis package: the ability to apply multiple sensitivity analyses to one sample, provision of tools to assess and revise user choices, and visualization tools.

Similarly, and extending these points, (Marelli and Sudret, 2014) outlined the limitations of currently available SA packages: current packages lack generality, are too complex, lack extendibility and the ability to use high-performance computing, are intrusive (require tight-coupling to the model code, rather than working as a black – or gray - box), lack portability between operating systems, and lack collaborative development.

More comprehensive packages released in recent years also include test functions and case studies for research and educational purposes. Documentation is more detailed and includes examples and tutorials. The packages surveyed did not appear to provide information about sampling and SA methods and which ones are most appropriate to given models. This emphasizes the importance of an active user community to share knowledge and update code.

The most comprehensive package in the survey, VARS-TOOL, provided users with global sensitivity analysis, VARS, sampling methods, SA with time-variance, factor grouping, and measures of the robustness and convergence of the SA, model emulation, visualization tools, and test functions. This leads to questions regarding what constitutes a thorough sensitivity analysis package, and how these packages can be best employed to service the various applications of SA: to identify the greatest contributing factors to model uncertainty, to assess the fidelity of the model to the underlying real-world situation, to understand the role and function of factors that are most influential on the model response, and to reduce models by identifying unresponsive or less influential factors in order to remove or constrain them (Razavi et al., 2019).

Comparison of surveyed pacakges.