

RESEARCH

Open Access



Can functional characteristics usefully define the cloud computing landscape and is the current reference model correct?

Neil Caithness* , Michel Drescher and David Wallom

Abstract

The NIST definition of cloud computing has been accepted by the majority of the community as the best available description to fully capture the variety of factors which determine how different stakeholders create, use or interact with cloud computing. With the breadth of the cloud computing landscape there is a need being expressed from within different cloud activities to consider how it may be best segmented so that the diversity might be more easily understood by the different stakeholders. The NIST definition considers four different deployment models (Private, Public, Hybrid, Community Cloud), three different service models (IaaS, PaaS, SaaS), and a number of characteristics (five in the final published version, but 13 in previous unpublished drafts). Exploring the definition further, this study aims to answer two questions: first, how can we use the affinity that different activities have with the definition's characteristics and second, how well does the definition describe the whole cloud ecosystem? We find that utilising a quantitative methodology shows a clustering of different cloud projects and activities that are technically aligned and therefore likely to benefit from interactions and shared learning, and that the final (short-list) definition is more robust than the draft (long-list) definition. Finally, we present a segmentation of the cloud landscape that we believe can best support a sharing of learning between projects in individual clusters.

Keywords: Cloud computing, Cloud ecosystem, NIST definition of cloud computing

Introduction

Since the emergence of cloud computing as a distinct paradigm within distributed computing, and as an important emerging market for ICT based services, there have been a number of efforts to support and encourage the adoption of cloud computing, as well as to foster a more geographically diverse cloud computing provider community. This has resulted in a large number of research and innovation projects receiving European Commission (EC) support over the past five years through the FP7 and H2020 programmes.

As part of the methodology to ensure success of supported projects there have been regular funder led attempts to bring projects together to share learning. Using the domain description as the key differentiator it was thought that synergistic clusters of projects similar

enough to share learning on both technical and social best practices would naturally emerge. Unfortunately, this has resulted in only limited appeal, as it has been unclear to participants exactly how the clusters are to be useful or effective as they are only superficially similar but differ widely in the cloud technologies and techniques in use.

Parallel to this European experience, in the US attempts to characterise the diverse landscape of cloud computing began at the National Institute of Standards and Technology (NIST). After years of development and 15 drafts, the final version of the cloud computing definition was published in September 2011 [5]. We describe this reference model more fully below but here we highlight a key feature—the model defines a limited set of functional characteristics that can be used to derive a quantitative description of the emerging cloud computing landscape. At a time when a large number of EC supported projects are maturing, we take the opportunity to make a quantitative assessment of their affinity to these functional characteristics and to derive a robust quantitative description of

*Correspondence: neil.caithness@oerc.ox.ac.uk
Oxford e-Research Centre, University of Oxford, 7 Keble Road, OX1 3QG
Oxford, UK

the cloud landscape. We identify clusters of projects that fully segment the landscape and provide a rational basis for enhancing shared learning.

Finally, appreciating the diversity of cloud computing activities is crucial to the process of deriving consistent and useful standards for cloud adoption and interoperability. We present a segmentation of the landscape of cloud computing that we believe can best inform this process.

ISO published cloud standards

The International Standards Organization and the International Electrotechnical Commission, together with the International Telecommunication Union, have recently released a new International Standard for cloud computing, *ISO/IEC 17788, Cloud computing—Overview and Vocabulary* and *ISO/IEC 17789, Cloud computing—Reference architecture*. These new standards provide definitions of common cloud computing terms, including those for cloud service categories such as Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). They also specify the terminology for cloud deployment models such as “public” and “private” cloud. Crucially, both new standards draw on previous developments by NIST, including *SP 800-145* [5] which has provided the primary dataset for this study. In light of these evolving developments we regard our present study as a timely appraisal of the foundational work.

Defining a methodology

Cloud computing resides in a complicated ecosystem of stakeholders with differing requirements and expectations. Even with a broad consensus that cloud computing is a general term describing anything that delivers hosted services over the Internet, interpretations of this vary widely and the field is subject to excessive hyperbole. Consequently, there is great latitude for interpretation on what exactly constitutes cloud computing and there are many approaches that might be adopted to gain insight into this dynamic and difficult to grasp ecosystem.

We require our approach to produce repeatable, insightful, and accurate information on the landscape and its segmentation. We therefore require the method to be evidence based rather than merely prescriptive. To this end, we define a methodology as follows:

1. Adopt the NIST characteristics of cloud computing as variables against which to make quantitative assessments.
2. Engage with project participants in the quantitative self-assessment of affinity to the characteristic variables.
3. Apply repeatable unsupervised machine learning techniques to these data as evidence-based

characterisation of the total cloud computing landscape.

4. Test for the robustness of the landscape as defined by the NIST characteristics.
5. Derive a clustering of cases (projects) along the complex dimensions of cloud computing and thereby facilitate the experience of shared learning.
6. Derive a segmentation of the landscape with respect to interactions among the variables (characteristics) to inform the process of standards development.

NIST defining characteristics of the cloud

The NIST model is the most commonly cited third party definition of cloud computing. As previously noted this definition took significant time and number of iterations before final publication, indicative of the difficulties associated with reaching a consensus on which even a restricted set of experts in the field could agree.

By consensus agreement [5, p.2]

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (...) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

The NIST model intends to capture this complexity in simple, understandable characteristics. The model is composed of five essential characteristics, three service models, and four deployment models. In earlier drafts of the definition NIST included a further eight common characteristics. These were dropped in the final published version. At the time of writing the earlier draft had been replaced in all on-line references that we could find. However, we were able to collect self-assessment data on the full set of 13 characteristics. We present analyses of both the NIST long- and short-lists.

Here we provide short explanations of each of the 13 characteristics:

Essential characteristics (1-5)

(Note that these descriptions of the five essentials are intended to be similar, but not necessarily identical to the NIST published descriptions. This was due to the need for brevity to give projects with which we engaged succinct and easy to understand and rate definitions.)

1 On-demand self-service

Consumers can log on to a website or use web services to access additional computing resources on demand, whenever they want, without human interference in the process.

2 Broad network access

Because they are web-based, you can access cloud-computing services from any internet-connected device.

With a web browser on a desktop machine, or even a thin client computer terminal, you can do any computing that the cloud resources provide.

3 Resource pooling

In multi-tenanted computing clouds the customers share a pool of computing resources with other customers, and these resources, which can be dynamically reallocated, may be hosted anywhere.

4 Rapid elasticity

Cloud computing enables computing resources or user accounts to be rapidly and elastically provisioned or released so that customers can scale their systems up and down at any time according to changing requirements.

5 Measured service

Cloud computing providers automatically monitor and record the resources used by customers or currently assigned to customers, which makes possible the pay-per-use billing model that is fundamental to the cloud-computing model.

Common characteristics (6-13)

6 Massive Scale

A cloud platform may, depending on the resources offered, provide individual users with access to large-scale or even massive-scale computing.

7 Homogeneity

In many situations, it is advantageous to both customers and providers to have essentially homogeneous systems at their disposal. Where requirements are particularly difficult or unusual, a cloud platform may be built out of non-homogeneous systems and components.

8 Virtualisation

Virtualisation of machines as software systems massively increases the scale of cloud resources that can be made available. Virtualisation is not an essential characteristic but it is becoming the only way that scale demands can be met by providers; customers generally don't care either way as the virtualisation is entirely transparent.

9 Low cost Software

If increased scale reduces per-unit, or per-use cost, then cloud-computing offers a drive towards lower-cost software. It is important to note that this may not be the case across all sectors and activities.

10 Resilient computing

In some sectors, continuous availability of computing with zero-downtime is crucial to the sectors requirements, for example, emergency and financial systems. In these

sectors, requirements for resilient, rather than just fail-safe computing will be the norm.

11 Geographic distribution

Some sectors have legal requirements that physical data stores are in particular geographical jurisdictions. This places certain restrictions on providers favouring a cloud-anywhere model. More commonly, the user is not concerned about location per se.

12 Service orientation

The design of the services that run and operate on the cloud frameworks are normally operated as services such they can take advantage of other factors that give resilience. This includes the ability to scale different components within the system depending on their load and capability.

13 Advanced security

There may be the capability to perform both system and network level security within the cloud system.

Data collection by quantitative self-assessment

Our final sample used in this analysis includes as many of the projects funded under the EC FP7-ICT-2013-10 calls 5, 7, 8 and 10 from which we were able gather data. We approached authoritative project representatives and asked them to self-asses their project's affinity to the 13 characteristics based on the descriptions given above on an ordinal scale from 0-9, with 0 indicating no affinity (the project is indifferent to considerations of this feature) and 9 indicating the strongest affinity (the project regards this feature to be crucially important).

We were able to compile responses from 37 projects (the SeaClouds project supplied two scorings independently from two different representatives). Table 1 gives a complete list of project names and self-assessed scores on the 13 variables.

Goals and techniques of analysis

Faced with a cases-by-variables data matrix (as in Table 1) the analyst seeks to summarise, simplify and explain. With high-dimensional data, there are a number of issues of interest:

- What are the relationships among the cases?
- What are the relationships among the variables?
- What insights may be gained from a summarized joint representation?
- Is the summary robust against possible variations and errors of case sampling?

Specifically for this analysis, the relationships among cases (cloud projects) lets us propose a clustering of

Table 1 Self-assessed scores of 38 cases (projects) on 13 variables

Project number	Project name	On demand self-service	Broad networked access	Resource pooling	Rapid elasticity	Measured service	Massive scale	Homogeneity	Virtualisation	Low cost software	Resilient computing	Geographic distribution	Service orientation	Advanced security
1	ARTIST	3	5	8	7	8	2	1	1	6	1	7	7	3
2	ASCETIC	7	2	5	7	9	7	3	5	7	8	7	6	2
3	BETaaS	7	8	6	7	6	4	3	4	2	7	6	7	5
4	BigFoot	9	4	9	9	1	9	1	9	1	6	4	4	1
5	BNCweb	7	4	4	3	2	4	5	3	5	2	1	7	2
6	Broker@Cloud	3	4	7	4	7	2	8	7	2	8	5	9	7
7	Catania Sci. Gateway	8	6	6	7	5	6	6	8	7	5	6	5	5
8	CELAR	8	4	7	9	5	2	4	7	5	9	6	7	3
9	CloudCatalyst	6	6	6	4	1	1	1	6	8	4	4	6	5
10	CloudLightning	9	7	8	7	6	9	5	5	6	5	5	5	3
11	CloudScale	9	9	6	9	6	9	3	6	7	1	1	9	1
12	CloudSpaces	9	9	9	9	9	9	7	9	9	9	7	9	9
13	CloudTeams	9	9	7	2	1	1	7	1	8	1	2	1	7
14	CloudWave	8	8	8	8	9	4	3	7	8	7	3	9	5
15	COMPOSE	7	4	6	7	6	2	4	7	7	6	2	9	4
16	DICE	2	2	7	8	8	6	2	6	8	9	7	6	6
17	Embassy Cloud	7	7	2	1	6	3	6	8	3	8	1	7	9
18	GEMMA	8	7	8	8	7	3	3	4	3	8	5	8	9
19	INPUT	7	5	5	8	6	4	2	8	2	5	8	5	2
20	IOStack	9	9	9	8	8	7	7	9	6	9	2	9	2
21	LEADS	9	8	2	4	3	7	7	8	8	8	5	9	3
22	Leicester	6	6	5	5	3	2	7	3	4	5	2	6	8
23	MCN	9	9	5	1	5	9	2	8	5	5	9	9	6
24	Mobizz	9	7	9	7	7	6	8	4	6	9	7	8	9
25	MODAClouds	8	4	5	9	9	9	1	1	8	8	8	9	1
26	OpenModeller	9	7	8	7	3	7	8	8	9	4	7	8	5
27	PaaSword	7	1	7	1	1	3	1	3	1	5	4	9	9
28	PANACEA	8	9	8	8	6	5	5	7	5	8	7	9	8
29	S-CASE	9	7	7	3	3	6	3	3	7	6	4	9	7
30	SeaClouds(1)	7	3	3	9	9	7	2	4	7	8	9	8	2
31	SeaClouds(2)	8	4	5	9	9	9	1	1	8	8	8	9	1
32	STORM CLOUDS	6	7	8	8	9	6	3	6	6	4	7	8	9
33	SUPERCLOUD	8	2	7	4	2	3	8	8	2	9	7	5	9
34	Texel	5	8	7	7	8	4	4	4	4	8	3	8	7
35	U-QASAR	5	7	6	7	7	2	2	5	4	4	4	7	6
36	Umea	5	3	3	2	2	2	4	6	4	3	3	7	7
37	Varberg	8	7	8	5	4	3	4	5	3	6	3	8	8
38	WeNMR	9	8	8	9	5	7	9	8	7	3	9	6	3

Scores are on an ordinal scale from 0-9, with 0 indicating no affinity and 9 indicating the strongest affinity. The first column shows the case number corresponding to the number used in the biplot and cluster diagrams

projects that are technically aligned and therefore likely to benefit from interactions and shared learning. The relationships among variables (NIST characteristics) lets us present a segmentation of the cloud landscape that we believe can inform where different projects may find examples of best practice or technology choices suitable in those type of projects. The joint representation informs projects about their position in the landscape, and informs standards developers about the relevant cloud activities when considering the landscape characteristics. Finally, measuring robustness for different partitions of the variables (all 13 characteristics; 5 essentials; 8 common) lets us reach conclusions about which of these partitions may be more meaningful and in that sense better.

Dimension reduction

For any high-dimensional dataset, the statistical techniques of dimension reduction are indispensable to the aim of summarizing, simplifying and explaining meaning contained in the dataset. Here we give a brief description of the linear algebra exploited by the family of multivariate dimension-reduction methods that includes principal component analysis (PCA), log-ratio analysis (LRA), correspondence analysis (CA), and various forms of discriminant analysis. All of these methods are basic decompositions (or factorizations) of a target matrix into left- and right-vector matrices representing respectively the cases (rows) and the variables (columns) of the original data matrix.

What distinguishes the various methods is the form of normalization and differential weighting of points, chosen depending on the type of data, which is applied before the decomposition into left- and right singular vector matrices.

For ordinal survey data such as ours, CA is the appropriate method [3]. The normalization in CA is the matrix of standardized residuals $T = D_r^{-1/2}(P - rc^T)D_c^{-1/2}$, where $P = N/n$ is the so-called correspondence matrix with N being the original data matrix and n its grand total. Row and column marginal totals of P are r and c respectively, and D_r and D_c are the diagonal matrices of these. Singular value decomposition (SVD) provides the appropriate factorization of T , with convenient properties, such that $T = U\Gamma V^T$. The right singular vectors V , are the contribution coordinates of the variables. A further transformation involving a scaling factor D_q , such that $F = D_q^{-1/2}U\Gamma$ defines the principal coordinates of the cases.

Visualization and interpretation

The biplot [2, 3] is a joint display of the two sets of points in V and F which, with the above transformation, can often be achieved on a common scale thereby avoiding the need for arbitrary independent scaling to make the biplot legible. The dimensions (Dim 1, Dim 2, ... Dim N, i.e. the

columns) of V and F are arranged in decreasing order of importance to the reduced-dimension solution. That means that a biplot of the first two dimensions is showing the most important relationships that can be represented on a planar 2-d plot. Cases are arranged in approximately Euclidean space so that proximity equates closely with similarity. Cases further from the origin have a stronger influence in the reduced-dimension solution; cases closer to the origin have a lesser influence. Variables are represented as vectors; the angular distance between vectors equates with correlation, and vector length again equates with relative contribution to the solution. One aspect of interpretation lies in identifying vectors with the same orientation, vectors at right angles, and vectors in opposing directions, and then identifying peripheral cases and how these are arranged with respect to the vectors. A second aspect of interpretation is to identify vectors aligned with the primary dominant dimension (Dim 1), and those aligned with the secondary dimension (Dim 2).

An important shortcoming of the biplot is that only two, or at most three, of the dimensions of the reduced-dimension solution can be visualised together. However, there may be more dimensions that are relevant for interpretation. The singular values of the SVD decomposition, in the description above, are the square roots of the eigenvalues; these indicate the relative importance of a dimension, i.e. its contribution to capturing information content in the analysis. When the eigenvalues are scaled so that their sum equals the number of variables, a conventional rule is to regard at least all those dimensions with eigenvalues > 1 as relevant. This is the conventional Kaiser-Guttman stopping rule [4, 6]. Scree plots below show eigenvalues with the Kaiser-Guttman reference line.

Addressing this shortcoming, we develop additional analytic displays in the form of hierarchical dendrograms based on the cosine distances between vectors in any number of dimensions. The hierarchical ordering of nodes in the tree then indicates closeness of orientation in the n -dimensional equivalent biplot.

Robustness testing

The robustness of a particular solution to possible errors and variation in sampling can be assessed by the statistical technique of bootstrapping [1]. The technique involves resampling rows from the data table, with replacement, to construct a replicate dataset of the same size as the original. The analysis is then repeated on the replicate, and the whole process is repeated a large number of times (typically 1000).

Computation

All analysis was performed on a standard desktop computer using Matlab Version 9 (R2016a) and custom

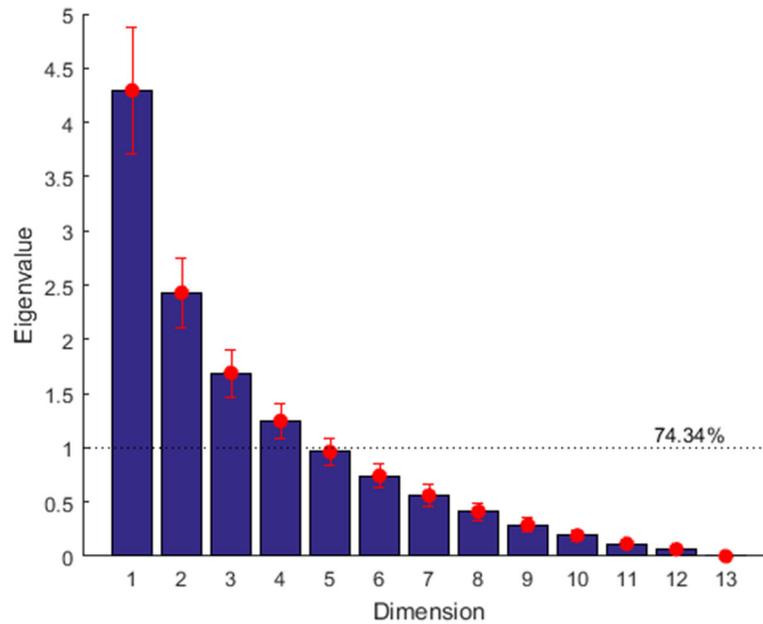


Fig. 2 Scree plot showing eigenvalues of the decomposition of the full 13-characteristic dataset. The dotted line shows the Kaiser-Guttman reference line for eigenvalues > 1 indicating that at least four of the 13 dimensions are relevant. These account for 74.34% of the total variance in the 13-dimensional dataset. Error bars show standard errors for 1000 bootstrap replicates of the CA analysis

survey of participants. Standard multivariate dimension reduction techniques provided the definition of the landscape which we visualise in biplots and cluster trees.

The robustness analysis of three partitions of the dataset (NIST long-list, short-list and residual-list) is decisive: any

interpretations drawn from the cloud landscape depicted in the analyses of the long- and residual-lists are essentially meaningless if applied to the general cloud ecosystem that these sample projects were drawn from. Only the NIST short-list of 5 essential characteristics leads

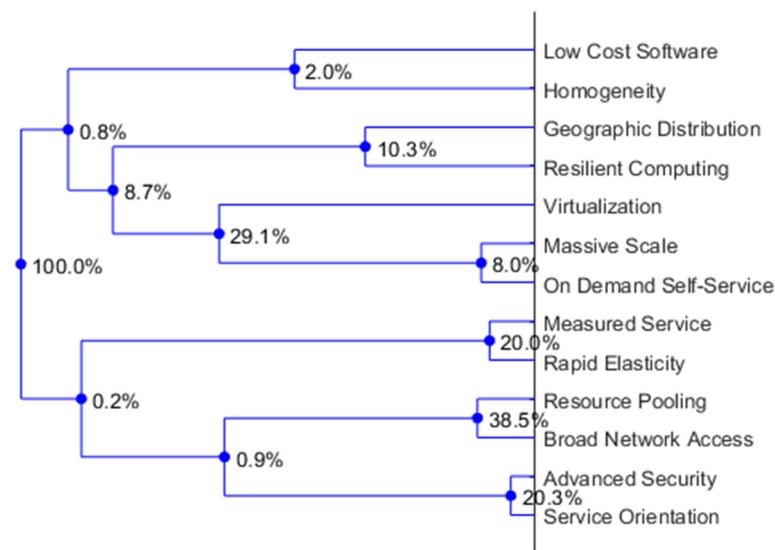


Fig. 3 Hierarchical dendrogram of variables in a 4-dimensional representation of the decomposition of the full 13-dimensional dataset. *Advanced Security* and *Massive Scale* are shown at opposite ends of the tree connected only by the deepest node, consistent with the view in the 2-d biplot. *Measured Service* and *Geographic Distribution*, shown very close together in 2-d, are in fact very distant when considering more relevant dimensions. Internal nodes are labelled with bootstrap percentages indicating the proportion of times in 1000 bootstrap replicates of the CA analysis that a node is observed. All values are < 50%, with many values much smaller, indicating that this solution is not robust to bootstrap resampling. i.e. no grouping shown here is strongly supported by evidence

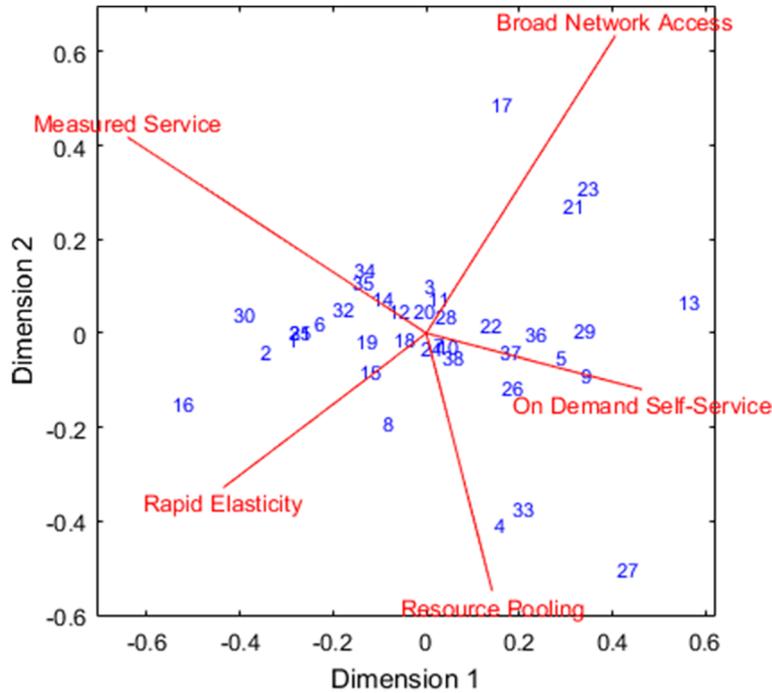


Fig. 4 2-dimensional biplot of the 5-dimensional partitioned dataset of the Essential Characteristics. Here *Measured Service* and *Broad Network Access* are slightly dominant in the solution, but perhaps insignificantly. *Broad Network Access* is driven largely by cases 17, 21 and 23, *Resource Pooling* by cases 27, 4 and 33, *Measured Service* and *Rapid Elasticity* jointly by a larger set of cases, in decreasing significance, 16, 2, 30, 1, 31, 25, 6 and 32. The cluster of cases 13, 29, 9, 5, 36, 37, 26 and 22, drives *On Demand Self-Service*. The remaining cases close to the centre form a distinct cluster of similarity, but do not contribute significantly to the 2-d solution

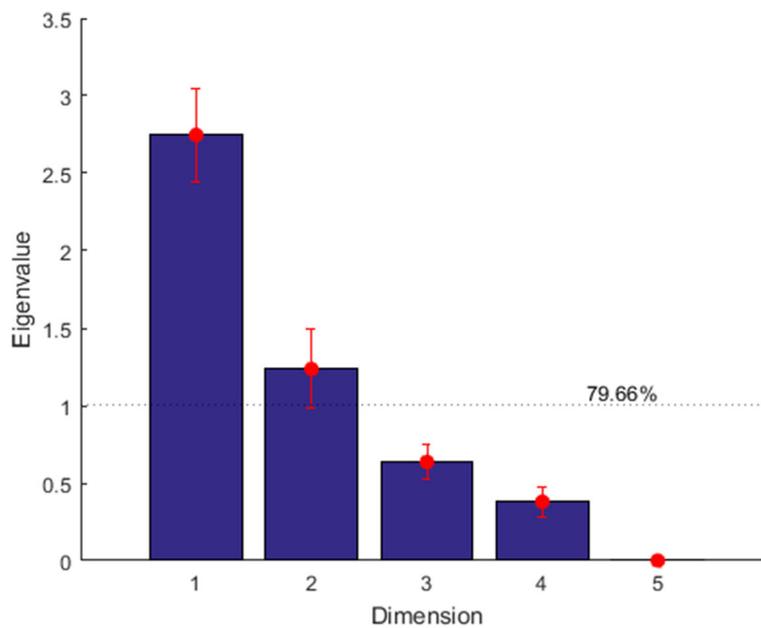


Fig. 5 Scree plot showing eigenvalues of the decomposition of the 5-dimensional partitioned dataset of the essential characteristics. The *dotted line* shows the Kaiser-Guttman reference line for eigenvalues > 1, indicating that at least two of the five dimensions are relevant. These account for 79.66% of the total variance in the 5-dimensional dataset. Error bars show standard errors for 1000 bootstrap replicates of the CA analysis

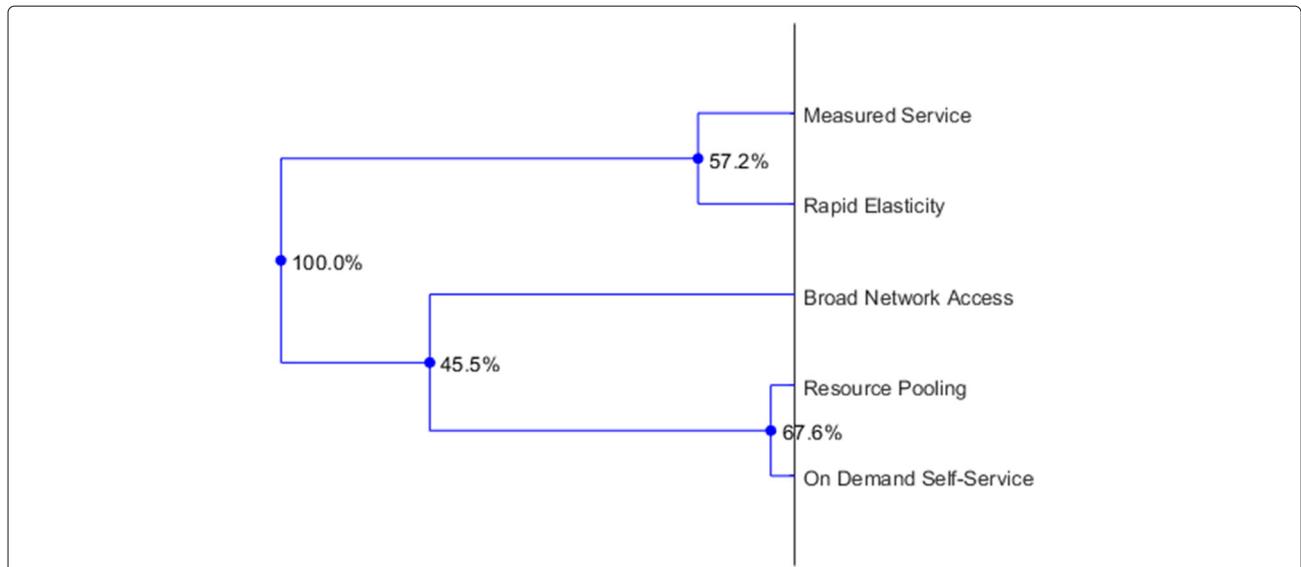


Fig. 6 Hierarchical dendrogram of variables in a 2-dimensional representation of the decomposition of the 5-dimensional partitioned dataset for the essential characteristics. Internal nodes are labelled with bootstrap percentages indicating the proportion of times in 1000 bootstrap replicates of the CA analysis that a node is observed. The two distal nodes with values >50% indicate that no other arrangement occurred more often in the bootstrap replicates, and that they are therefore strongly supported. The association of *Broad Network Access* with either distal node is then the only remaining degree of freedom, and occurs roughly half of the time with each (45.5% vs. 54.5%). Looking again at the biplot in Fig. 4 we see that this would result from only slight alterations to the angle of the vector towards either *On Demand Self-Service*, or towards *Measured Service*. Overall, this is a robust result and the two association of (i) *Resource Pooling* with *On Demand Self-Service*, and (ii) *Measured Service* with *Rapid Elasticity* are both stable and well supported by the evidence

to a stable and robust depiction of the landscape and general interpretations should be made only against this depiction.

Figure 4 (biplot) and Fig. 7 (cluster tree) provide the interpretive mechanisms. Five groups of projects labelled I-V in Fig. 7 are a useful focus for discussion, though clusters at any level of the tree are equally meaningful.

Cluster I

- 23 MCN
- 21 LEADS
- 17 Embassy Cloud

These are the primary drivers of *Broad Network Access*, which is also one of the two dominant features of the landscape.

Cluster II

- 13 CloudTeams
- 26 OpenModeller
- 37 Varberg
- 36 Umea
- 22 Leicester
- 29 S-CASE
- 9 CloudCatalyst
- 5 BNCweb

The primary drivers of *On Demand Self Service*, with CloudTeams being the most significant to this axis, and Leicester being the least.

Cluster III

- 27 PaaSsword
- 33 SUPERCLOUD
- 4 BigFoot

The primary drivers of *Resource Pooling*, with PaaSsword being the most significant.

Cluster IV

- 35 U-QASAR
- 34 Texel
- 19 INPUT
- 15 COMPOSE
- 18 GEMMA
- 14 CloudWave
- 12 CloudSpaces
- 8 CELAR
- 38 WeNMR
- 10 CloudLightning
- 24 Mobizz
- 7 Catania Sci. Gateway
- 28 PANACEA
- 20 IOStack

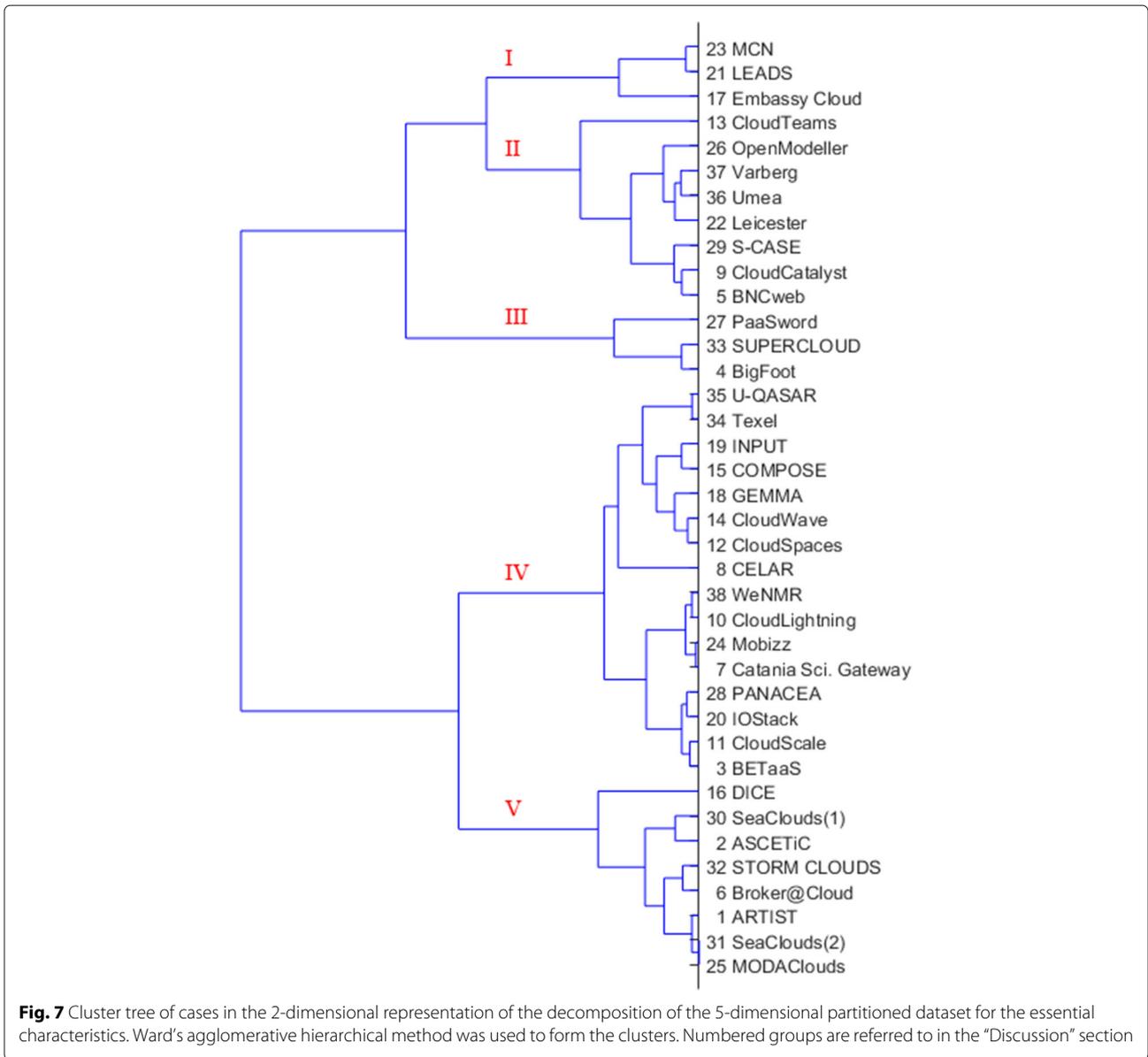


Fig. 7 Cluster tree of cases in the 2-dimensional representation of the decomposition of the 5-dimensional partitioned dataset for the essential characteristics. Ward’s agglomerative hierarchical method was used to form the clusters. Numbered groups are referred to in the “Discussion” section

- 11 CloudScale
- 3 BETaaS

The large central cluster not distinguished by a tendency towards any of the NIST characteristics over any other.

Cluster V

- 16 DICE
- 30 SeaClouds(1)
- 2 ASCETiC
- 32 STORM CLOUDS
- 6 Broker@Cloud
- 1 ARTIST
- 31 SeaClouds(2)
- 25 MODAClouds

This cluster subtends the two remaining characteristics, *Measured Service* and *Rapid Elasticity*, with DICE being the most significant and leaning towards *Rapid Elasticity*. As a final observation we note that although *Measured Service* is one of the dominant features of the landscape, no projects identify significantly with this feature.

Finally, we suggest that this interpretation of project clusters can provide the basis for enhanced shared learning among projects that are technically aligned on the axes of cloud characteristics, and that the further development of standards for cloud implementation and adoption can benefit from this depiction of the landscape and the association shown between projects and characteristics.

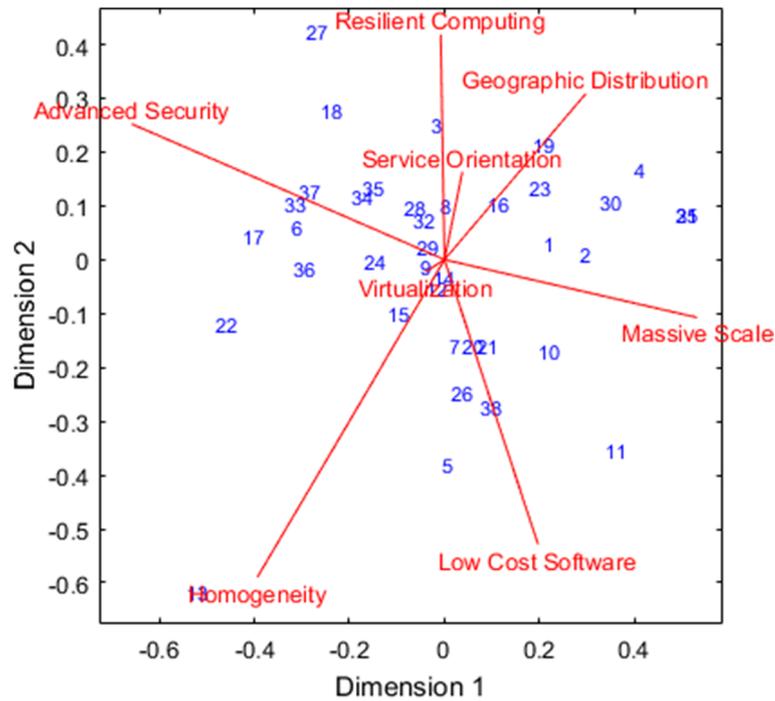


Fig. 8 2-dimensional biplot of the 8-dimensional partitioned dataset of the common characteristics. *Service Orientation* and *Virtualisation* appear almost insignificant to this solution. There is a roughly even spread of cases, though similar inferences could be made as illustrated in Fig. 1

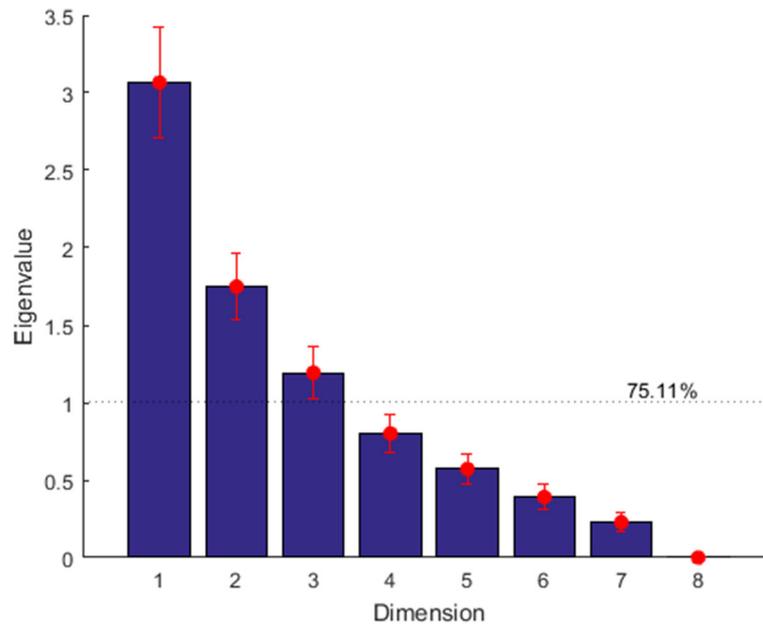
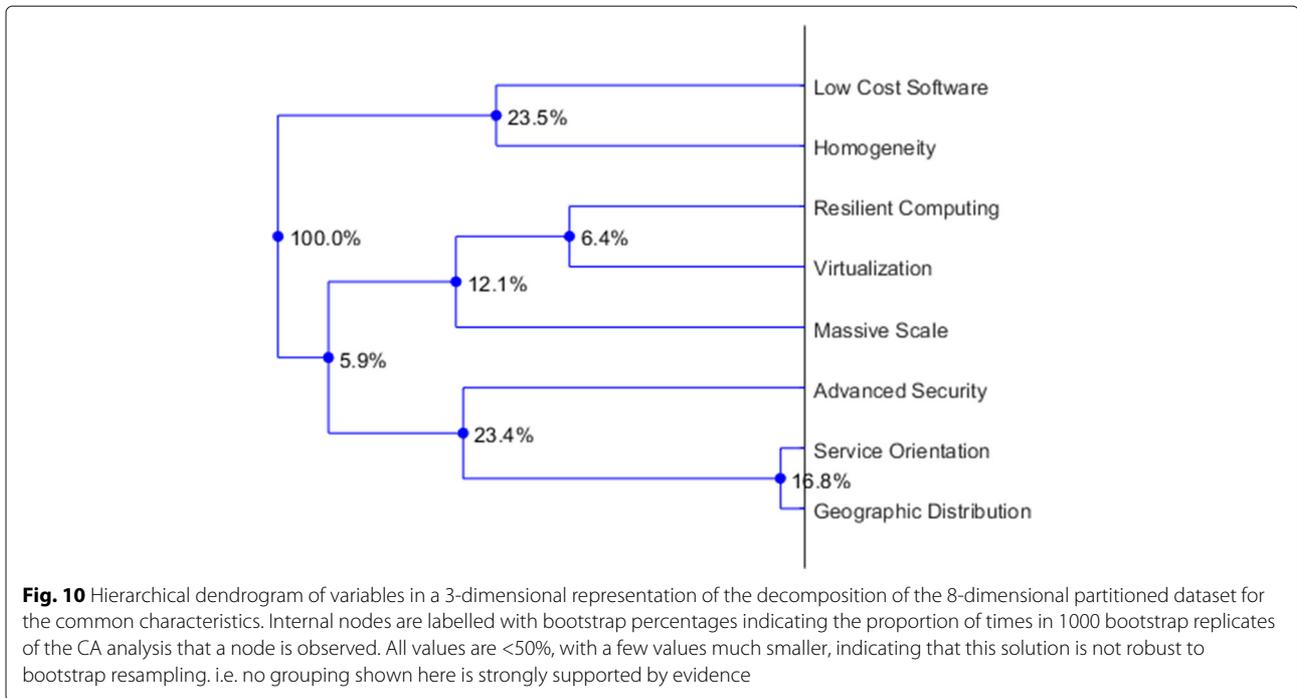


Fig. 9 Scree plot showing eigenvalues of the decomposition of the 8-dimensional partitioned dataset for the common characteristics. The *dotted line* shows the Kaiser-Guttman reference line for eigenvalues > 1, indicating that at least three of the eight dimensions are relevant. These account for 75.11% of the total variance in the 8-dimensional dataset. Error bars show standard errors for 1000 bootstrap replicates of the CA analysis



Responses to the methodology and future work

In this paper, we have presented a methodology for characterising the landscape of cloud computing based on the set of NIST defining features. The same methodology identifies the location of a project or cloud enterprise within the landscape. Taken together, the resulting biplot and the cluster tree offer rich interpretive tools in defining the current cloud computing landscape.

Presenting these results to partners in the supporting projects, and to the wider circle of EC-funded cloud related projects, participants accepted the methodology as sound and applicable, and they generally regarded the characterisation of the landscape as useful. However, several observations are noteworthy.

- In draft form, *SP 800-145* defines characteristics of cloud computing only, however, many more IT service characteristics are applicable that are not specific to cloud computing.
- Several of the characteristics in the draft form subsume important aspects under a single term that is too general to be meaningful, such as *Privacy* and *Data Protection* being subsumed under the term *Advanced Security*.
- Despite *SP 800-145* having been published over five years ago in 2011, participants disagree on interpretation of certain characteristics. *Measured Service*, for example, is frequently misinterpreted as describing the monitoring of a cloud service system.

Several avenues of future work have emerged from this activity.

- An automated web-hosted tool that allows the self-assessment of new cloud enterprises and shows their location within the existing landscape would be informative to new enterprises.
- Exploring the impact of reducing the set of characteristics from 13 to five may be instructive. For instance, does the larger set define a much finer grained landscape than the reduced set, or is the information content of the reduced set sufficient?
- Does expanding the set of characteristics to include more applicable IT service characteristics improve the resolution of the landscape, or does this perhaps significantly change the orientation of the existing characteristics?

Acknowledgements

The authors thank two anonymous reviewers for helpful comments and suggestions.

Funding

This work was supported by the EC Framework 7 CloudWatch project (EC grant no. 610994).

Authors' contributions

DW proposed a landscape segmentation analysis using the NIST model characteristics as variables for a quantitative study and drafted the descriptive definitions. MD wrote the first draft of the paper and interpreted appropriate segmentations of the cluster tree. NC devised the methodology, liaised with projects on data collection, implemented the software, performed all computation, and wrote subsequent drafts of the paper with significant

contributions from the co-authors. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Received: 22 December 2016 Accepted: 23 May 2017

Published online: 05 June 2017

References

1. Efron B, Tibshirani RJ (1994) An Introduction to the Bootstrap. Chapman and Hall/CRC, New York
2. Greenacre M (2010) Biplots in Practice. BBVA Foundation, Madrid
3. Greenacre M (2013) Contribution biplots. *J Comput Graph Stat* 22:107–122
4. Jackson DA (1993) Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology* 74(8):2204–2214
5. Mell P, Grance T (2011) The NIST definition of cloud computing. US National Institute of Standards and Technology. Available: <http://dx.doi.org/10.6028/NIST.SP.800-145>
6. Peres-Neto PR, Jackson DA, Somers KM (2005) How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Comput Stat Data Anal* 49(4):974–997

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
