Ann Oper Res
https://doi.org/10.1007/s10479-018-2795-1

CrossMark

Coherent quality management for big data systems: a dynamic approach for stochastic time consistency

Yi-Ting Chen[1,2] · Edward W. Sun[3] · Yi-Bing Lin[2]

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract Big data systems for reinforcement learning have often exhibited problems (e.g., failures or errors) when their components involve stochastic nature with the continuous control actions of reliability and quality. The complexity of big data systems and their stochastic features raise the challenge of uncertainty. This article proposes a dynamic coherent quality measure focusing on an axiomatic framework by characterizing the probability of critical errors that can be used to evaluate if the conveyed information of big data interacts efficiently with the integrated system (i.e., system of systems) to achieve desired performance. Herein, we consider two new measures that compute the higher-than-expected error,—that is, the tail error and its conditional expectation of the excessive error (conditional tail error)—as a quality measure of a big data system. We illustrate several properties (that suffice stochastic time-invariance) of the proposed dynamic coherent quality measure for a big data system. We apply the proposed measures in an empirical study with three wavelet-based big data systems in monitoring and forecasting electricity demand to conduct the reliability and quality management in terms of minimizing decision-making errors. Performance of using our approach in the assessment illustrates its superiority and confirms the efficiency and robustness of the proposed method.

Keywords Big data · Dynamic coherent measure · Optimal decision · Quality management · Time consistency

JEL Classification C02 · C10 · C63

✉ Edward W. Sun
edward.sun@bem.edu

1 School of Business Informatics and Mathematics, University of Mannheim, Mannheim, Germany

2 School of Computer Science, National Chiao Tung University (NCTU), Hsinchu, Taiwan

3 KEDGE Business School, 680 Cours de la Libèration, 33405 Talance Cedex, France

uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

1 Introduction

Big data have become a torrent, flowing into every area of business and service along with the integration of information and communication technology and the increasingly development of decentralized and self-organizing infrastructures (see Sun et al. 2015, for example). A big data system (BDS), which in our article refers to the big data-initiated system of systems, is an incorporation of a task-oriented or dedicated system that interoperates with big data resources, integrates their proprietary infrastructural components, and improves their analytic capabilities to create a new and more sophisticated system for more functional and superior performance. For example, the Internet of Things (IoT), as described by Deichmann et al. (2015) is a BDS, and its network's physical objects through the application of embedded sensors, actuators, and other devices can collect or transmit data about the objects. BDS generates an enormous amount of data through devices that are networked through computer systems.

Challenges to BDS are therefore unavoidable with the tremendously ever-increasing scale and complexity of systems. Hazen et al. (2014) address the data quality problem in the context of supply chain management and propose methods for monitoring and controlling data quality. The fundamental challenge of BDS is that there exists an increasing number of software and hardware failures when the scale of an application increases. Pham (2006) discusses the system software reliability in theory and practice. Component failure is the norm rather than exception, and applications must be designed to be resilient to failures and diligently handle them to ensure continued operations. The second challenge is that complexity increases with an increase in scale. There are more component interactions, increasingly unpredictable requests on the processing of each component, and greater competition for shared resources among interconnected components. This inherent complexity and non-deterministic behavior makes diagnosing aberrant behavior an immense challenge. Identifying whether the interruption is caused by the processing implementation itself or unexpected interactions with other components turns out to be rather ambiguous. The third challenge is that scale makes a thorough testing of big data applications before deployment impractical and infeasible.

BDS pools all resources and capabilities together and offers more functionality and performance with sophisticated technologies to ensure the efficiency of such an integrated system. Therefore, the efficiency of BDS should be higher than the merged efficiency of all constituent systems. Deichmann et al. (2015) point out that the replacement cycle for networked data may be longer than the innovation cycle for the algorithms embedded within those systems. BDS therefore should consider ways to upgrade its computational capabilities to enable continuous delivery of data (information) updates. Modular designs are required so the system can refresh discrete components (e.g., data) of a BDS on a rolling basis without having to upgrade the whole database. To pursue a continuous-delivery model of information or decision, BDS must be enabled to review its computational processes and be focused simultaneously on supporting data flows that must be updated quickly and frequently, such as auto-maintenance. In addition, BDS has to ensure speed and fidelity in computation and stability and security in operation. Networking efficiency is huge, but there is a lot of uncertainty over it. Questions thus arise about how to accurately assess the performance of BDS, and how to build a technology stack[1] to support it.

Quality measures quantify a system's noncompliance with the desirable or intended standard or objective. Such noncompliance may occur for a process, outcome, component,

[1] These are layers of hardware, firmware, software applications, operating platforms, and networks that make up IT architecture.

Uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

structure, product, service, procedure, or system, and its occurrence usually causes loss. Enlightened by the literature of managing the extreme events and disasters, in this article we work on an axiomatic framework of new measures (based on the tail distribution of critical defect or error) to address how to conduct optimal quality management for BDS with respect to a dynamic coherent measurement that has been well established particularly for managing risk or loss—see, for example, Riedel (2004), Artzner et al. (2007), Sun et al. (2009), Bion-Nadal (2009) and Chun et al. (2012) as well as references therein. Conventional quality measures focus on quantizing either the central moment (i.e., the mean) or the extreme value (e.g., six sigma) of error (or failure). The newly proposed measure focuses on a specific range of distribution instead of a specific point, which modulates the overall exposure of defects for quality management.

Conventional measurement based on average performance such as mean absolute error (MAE) is particularly susceptible to the influence of outliers. When we conduct system engineering, systems have their individual uncertainties that lead to error propagation that will distort the overall efficiency. When these individual uncertainties are correlated with an unknown pattern, simple aggregating a central tendency measure gives rise to erroneous judgment. The six sigma approach that investigates the extremum[2] has received considerable attention in the quality literature and general business and management literature. However, this approach has shown various deficiencies such as rigidity, inadequate for complex system, and non-subadditivity, particularly, not suitable for an integrated system or system of dynamic systems. However, Baucells and Borgonovo (2013) and Sun et al. (2007) suggested the Kolmogorov–Smirnov (KS) distance, Cramér–von Mises (CVM) distance, and Kuiper (K) distance as the metrics for evaluation. These measures are based on the probability distribution of malfunction and only provide information for relative comparison of extremum. They fail to identify the absolute difference when we need to know the substantiality of difference. In order to overcome these shortcomings, we propose a dynamic measurement applying tail error (TE) and its derivation, conditional tail error (CTE), to quantify the higher-than-expected failure (or critical error) of a BDS.

Tail error (TE) describes the performance of a system in a worst-case scenario or the critical error for a really bad performance with a given confidence level. TE measures an error below a given probability α that can be treated as a percentage of all errors. α is then the probability such that an error greater than TE is less than or equal to α, whereas an error less than TE is less than or equal to $1-\alpha$. For example, one can state that it is 95% (i.e., $\alpha = 5\%$) sure for a given system processing that the highest defect rate is no more than 15% (which is the tail error that can be tolerated). TE can also be technically defined as a higher-than-expected error of an outcome moving more than a certain tolerance level,—for example, a latency with three standard deviations away from the mean, or even six standard deviations away from the mean (i.e., the six sigma). For mere mortals, it has come to signify any big downward move in dependability. Because an expectation is fairly subjective, a decision maker can appropriately adjust it with respect to different severities.

When we assess (e.g., zoom in) tail-end (far distance from zero) along with the error (or failure) distribution, the conditional tail error (CTE) is a type of sensitivity measure derived from TE describing the shape of the error (or failure) distribution for the far-end tail departing from TE. CTE also quantifies the likelihood (at a specific confidence level) that a specific error will exceed the tail error (i.e., the worst case for a given confidence level), which is mathematically derived by taking a weighted average between the tail error and the errors

[2] A six sigma process is one in which 99.99966% of all opportunities are statistically expected to be free of defects (i.e., 3.4 defective features per million opportunities).

uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

exceeding it. CTE evaluates the value of an error in a conservative way, focusing on the most severe (i.e., intolerable) outcomes above the tail error. The dynamic coherent version of CTE is the upper bound of the exponentially decreasing bounds for TE and CTE. CTE is convex and can be additive for an independent system. A preternatural character of them is to quantitate the density of errors by choosing different α values (or percentages). Therefore, a subjective tolerance (determined by α) of error can be predetermined by a specific utility function of the decision maker. Intrinsic features of the conditional probability guarantee its consistency and coherence when applying it to dynamic quality management throughout the entire process. We show several properties of these measures in our study and explain how to compute them for quality management of a BDS.

We organize the article as follows. Section 2 introduces some background information focusing on BDS and quality management. We describe the prevailing paradigm for the quality management of a big data system. We then introduce the methodology in more details and illustrate several properties of the proposed framework. Section 3 presents three big data systems (i.e., SOWDA, GOWDA, and WRASA) based on wavelet algorithms. Section 4 conducts an empirical study to evaluate these three systems' performances on the big electricity demand data from France through analysis and forecasting. It also provides a discussion with simulation for several special cases that have not been observed in the empirical study. Our empirical and simulation results confirm the efficiency and robustness of the method proposed herein. We summarize our conclusions in Sect. 5.

2 Quality management for big data system

Optimal quality management refers to the methodologies, strategies, technologies, systems, and applications that analyze critical data to help an institution or individual better understand the viability, stability, and performance of systems with timely beneficial decisions in order to maintain a desired level of excellence. In addition to the data processing and analytical technologies, quality management for BDS assesses macroscopic and microscopic aspects of system processing and then combines all relevant information to conduct various high-impact operations to ensure the system's consistency. Therefore, the foundation of quality management for BDS is how to efficiently evaluate the interoperations of all components dealing with the big data through robust methodologies (see Sun et al. 2015, for example). In this section we first classify BDS—that is, the fundamentals for the system of systems engineering. We then discuss quality lifecycle management (QLM) for systematic quality management. We propose the framework of dynamic coherent quality management, which is a continuous process to consistently identify critical errors. We then set up the quantitative framework of metrics to evaluate the decision-making quality of BDS in terms of error control.

2.1 Big data system

The integration and assessment of a broad spectrum of data—where such data usually illustrate the typical "6Vs" (i.e., variety, velocity, volume, veracity, viability, and value)—affect a big data system. Sun et al. (2015) summarize four features of big data as follows: (1) autonomous and heterogeneous resource, (2) diverse dimensionality, (3) unconventional processing, and (4) dynamic complexity. Monitoring and controlling data quality turn out to be more critical (see Hazen et al. 2014, for example). In our article we refer to the big data

uncorrected proof

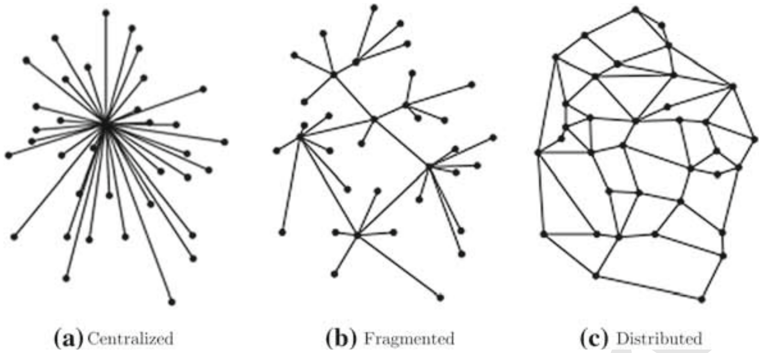


Fig. 1 Fundamental architectures of a big data system (BDS). Dots indicate components (processors and terminals) and edges are links. The component could be a system or a unit. **a** Centralized BDS that has a central location, using terminals that are attached to the central. **b** Fragmented (or decentralized) BDS that allocates one central operator to many local clusters that communicate with associated processors (or terminals). **c** Distributed BDS that spreads out computer programming and data on many equally-weighted processors (or terminals) that directly communicate with each other

initiated a system of systems as BDS, which is an incorporation of task-oriented or dedicated systems that interoperate their big data resources, integrate their proprietary infrastructural components, and improve their analytic capabilities to create a new and more sophisticated system for more functional and superior performance. A complete BDS contains (1) data engineering that intends to design, develop, integrate data from various resources, and run ETL (extract, transform and load) on top of big datasets and (2) data science that applies data mining, machine learning, and statistics approaches to discovery knowledge and develop intelligence (such as decision making).

When integrating all resources and capabilities together to a BDS for more functionality and performance, such a processing is called system of systems engineering (SoSE) (see Keating and Katina 2011, for example). SoSE for BDS integrates various, distributed, and autonomous systems to a large sophisticated system that pools interacting, interrelated, and interdependent components as an incorporation. We illustrate three fundamental architectures of BDS with Fig. 1 centralized, fragmented, and distributed.

Centralized BDS has a central operator that serves as the acting agent for all communications with associated processors (or terminals). Fragmented (or decentralized) BDS allocates one central operator to many local clusters that communicate with associated processors (or terminals). Distributed BDS spreads out computer programming and data on many equally weighted processors (or terminals) that directly communicate with each other. For more details about data center network topologies, see Liu et al. (2013) and references therein.

We can identify five properties of BDS following “Maier’s criteria” (see Maier 1998): (1) operational independence, i.e., each subsystem of BDS is independent and achieves its purposes by itself; (2) managerial independence, that is, each subsystem of BDS is managed in large part for its own purposes rather than the purposes of the whole; (3) heterogeneity, that is, different technologies and implementation media are involved and a large geographic extent will be considered; (4) emergent behavior, that is, BDS has capabilities and properties in which the interaction of BDS inevitably results in behaviors that are not predictable in advance for the component systems; and (5) evolutionary development, that is, a BDS evolves with time and experience. This classification shares some joint properties that can be used to characterize a complex system,—for example, the last three properties of heterogeneity, emergence, and evolution.

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

2.2 Systematic quality management

Modeling a system's featuring quality helps characterize the availability and reliability of the underlying system. The unavailability and unreliability of a system presents an overshoot with respect to its asymptotic metrics of performance with the hazard that characterizes the failure rate associated with design inadequacy. Levels of uncertainty in modeling and control are likely to be significantly higher with BDS. Greater attention must be paid to the stochastic aspects of learning and identification. Uncertainty and risk must be rigorously managed.

These arguments push for a renewed mechanism on system monitoring, fault detection and diagnosis, and fault-tolerant control that include protective redundancies at the hardware and software level (see Shooman 2002). Rigorous, scalable, and sophisticated algorithms shall be applied for assuring the quality, reliability, and safety of the BDS. Achieving consistent quality in the intelligent decision of the BDS is a multidimensional challenge with increasingly global data flows, a widespread network of subsystems, technical complexity, and a highly demanding computational efficiency.

To be most effective, quality management should be considered as a continuous improvement program with learning as its core. Wu and Zhang (2013) point out that exploitative-oriented and explorative-oriented quality management need to consistently ensure quality control on the operation lifecycle throughout the entire process. Parast and Adams (2012) highlight two theoretical perspectives (i.e., convergence theory and institutional theory) of quality management over time. Both of them emphasize that quality management evolves over time by using cross-sectional, adaptive, flexible, and collaborative methods so that the quality information obtained in one lifecycle stage can be transferred to relevant processes in other lifecycle stages. Contingency theory also recommends that the operation of quality control be continually adapted (see Agarwal et al. 2013, for example), and quality information must therefore be highly available throughout the whole system to ensure that any and all decisions that may require quality data or affect performance quality are informed in a timely, efficient, and accurate fashion.

Quality lifecycle management (QLM)[3] is system-wide, cross-functional solution to ensure that processing performance, reliability, and safety are aligned with the requirements set for them over the course of the system performance. Mellat-Parst and Digman (2008) extend the concept of quality beyond the scope of a firm by providing a network perspective of quality. QLM is used to build quality, reliability, and risk planning into every part of the processing lifecycle by aligning functional needs with performance requirements, ensuring that these requirements are met by specific characteristics and tracking these characteristics systematically throughout design, testing, development, and operation to ensure the requirements are met at every lifecycle stage.

Outputs from each lifecycle stage, including analysis results, performance failures, corrective actions, systemic calibrations, and algorithmic verifications, are compiled within a single database platform using QLM. They are made accessible to other relevant lifecycle stages using automated processes. This guarantees the continuous improvement of systemic performance over the course of development and further improvement. QLM links together the quality, reliability, and safety activities that take place across every stage of the product development lifecycle. Through QLM, one lifecycle stage informs the next, and the feedback from each stage is automatically fed into other stages to which it relates, we then obtain a unified and holistic view of overall quality of performance.

[3] PTC white paper. PTC is a global provider of technology platforms and solutions that transform how companies create, operate, and service the "things" in the Internet of Things (IoT). See www.ptc.com.

Uncorrected proof

2.3 The methodology

In this section, we describe the axiomatic framework for QLM with respect to the dynamic coherent error measures. Based on the Kullback–Leibler distance, we propose two concepts—that is, the tail error (TE) and conditional tail error (CTE)—and their dynamic setting to guarantee the feature of time consistency. These two error measures can be applied dynamically to control performance errors at each stage of BDS in order to conduct QLM.

2.3.1 Notation and definitions

The uncertainty of future states shall be represented by a probability space $(\Omega, \mathcal{F}, P)$ that is, the domain of all random variables. At time t a big data system is employed to make a decision about future scenario after a given time horizon Δ, for example, to predict price change from t to $t + \Delta$. We denote a quantitative measure of the future scenario at time t by $V(t)$ that is a random variable and observable through time. The difference between the predicted scenario at time t and its true observation at time $t + \Delta$ is then measured by $V(t + \Delta) - V(t)$. If we denote $\widetilde{V}(\cdot)$ as the prediction and $V(\cdot)$ the observation, then the Kullback–Leibler distance or divergence of the prediction error is given by the following definition:

Definition 1 For given probability density functions $V(x)$ and $\widetilde{V}(x)$ on an open set $\mathcal{A} \in \Re^N$, the Kullback–Leibler distance or divergence of $\widetilde{V}(x)$ from $V(x)$ is defined by

$$\mathcal{K}(V \| \widetilde{V}) = \int_{\mathcal{A}} V(x) \log \frac{V(x)}{\widetilde{V}(x)} dx. \tag{1}$$

Axiom 1 *Assume that $V(x)$ and $\widetilde{V}(x)$ are continuous on an open set $\mathcal{A}$, then for arbitrary $V(x)$ and $\widetilde{V}(x)$, $\mathcal{K}(V \| \widetilde{V}) \geq 0$; and $\mathcal{K}(V \| \widetilde{V}) = 0$ if and only if $V(x) = \widetilde{V}(x)$ for any $x \in \mathcal{A}$.*

Because $\widetilde{V}(x)$ and $V(x)$ are density functions, we have $\int \widetilde{V}(x)\mathrm{d}x = 1$ and $\int V(x)\mathrm{d}x = 1$, which shows Axiom 1.

Axiom 2 *Consider a set $\mathcal{K}$ of real-valued random variables. A functional $\rho : \mathcal{D} \to \mathbb{R}$ is referred to be a consistent error measure for $\mathcal{D}$ if it satisfies all of the following axioms:*

(1) *For all $\mathcal{K}_1, \mathcal{K}_2, \in \mathcal{D}$ such that $\mathcal{K}_2 \geq \mathcal{K}_1$ a.s., we have $\rho(\mathcal{K}_1) \leq \rho(\mathcal{K}_2)$.*
(2) *$\forall \mathcal{K}_1, \mathcal{K}_2, \in \mathcal{D}$ such that $\mathcal{K}_1 + \mathcal{K}_2 \in \mathcal{D}$, we have $\rho(\mathcal{K}_1 + \mathcal{K}_2) \leq \rho(\mathcal{K}_1) + \rho(\mathcal{K}_2)$; for $\mathcal{K}_1$ and $\mathcal{K}_2$ are comonotonic random variables, that is, for every $\omega_1, \omega_2 \in \Omega$: $(\mathcal{K}_1(\omega_2) - \mathcal{K}_1(\omega_1))(\mathcal{K}_2(\omega_2) - \mathcal{K}_2(\omega_1)) \geq 0$, we have $\rho(\mathcal{K}_1 + \mathcal{K}_2) = \rho(\mathcal{K}_1) + \rho(\mathcal{K}_2)$.*
(3) *$\forall \mathcal{K} \in \mathcal{D}$ and $\gamma \in \Re_+$ such that $\gamma\mathcal{K} \in \mathcal{D}$, we have $\rho(\gamma\mathcal{K}) \leq \gamma\rho(\mathcal{K})$.*
(4) *$\forall\, \mathcal{K} \in \mathcal{D}$ and $a \in \Re$ such that $\mathcal{K} + a \in \mathcal{D}$, we have $\rho(X + a) = \rho(X) + a$.*
(5) *$\forall\, \mathcal{K}_1, \mathcal{K}_2, \in \mathcal{D}$ with cumulative distribution functions $F_{\mathcal{K}_1}$ and $F_{\mathcal{K}_2}$ respectively, if $F_{\mathcal{K}_1} = F_{\mathcal{K}_2}$ then $\rho(\mathcal{K}_1) = \rho(\mathcal{K}_2)$.*

This axiom is motivated by the coherent measure proposed by Artzner et al. (2007). Here, we modify the axiomatic framework of consistency in describing the error function. By Axiom 2-(1), a monotonicity of the error function is proposed: the larger the Kullback–Leibler distance is, the larger the error metric will be. Axiom 2-(2) says that if we combine two tasks, then the total error of them is not greater than the sum of the error associated with each of them, if these two tasks are independent. When these two tasks are dependent, the total error of them is equal to the sum of the error associated with each of them. It is also called subadditivity in

Uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

risk management and also is applied in systematic computing such that an integration does not create any extra error. Axiom 2-(3) can be derived if Axiom 2-(2) holds, and it reflects positive homogeneity. It shows that an error reduction requires that processing uncertainty should linearly influence task loans. Axiom 2-(4) shows that the error will not be altered by adding or subtracting a deterministic quantity a to a task. Axiom 2-(5) states the probabilistic equivalence that is derived directly from Axiom 1.

2.3.2 Dynamic error measures

A dynamic error measure induces for each operation a quality control process describing the error associated with the operation over the time. O'Neill et al. (2015) point out that quality management helps maintain the inherent characteristics (distinguishing features) in order to fulfill certain requirements. The inherence of quality requires a dynamic measure to conduct continuous processing of error assessment in the rapidly changing circumstances of an operational system. During the dynamic phase, the decision-making process involves investigating the errors generated by the big data systems, selecting an appropriate response (system of work), and making a judgment on whether the errors indicate a systematic failure. How to interrelate quality control process over different periods of time is important for the dynamic measure. Consequently, a dynamic error measure should take into account the information available at the time of assessment and update the new information continuously over time. Time consistency then plays a crucial role in evaluating the performance of dynamic measures, particularly for big data systems that run everlastingly before stopping. Therefore, the quality measure for errors shall preserve the property of consistency.

Bion-Nadal (2008) defines a dynamic risk measures based on a continuous time filtered probability space as $(\Omega, \mathcal{F}_\infty, (\mathcal{F}_t)_{t\in\Re^+}, P)$ where the filtration $(\mathcal{F}_t)_{t\in\Re^+}$ is right continuous. Similarly, following the Axiom 2, we can extend it to a dynamic error measure following Bion-Nadal (2008), Chen et al. (2017), and references therein.

Definition 2 For any stopping time τ, we define the σ-algebra $\mathcal{F}_\tau$ by $\mathcal{F}_\tau = \{\mathcal{K} \in \mathcal{F}_\infty \mid \forall t \in \Re^+ : \mathcal{K} \cap \{\tau \le t\} \in \mathcal{F}_t\}$. Then $L^\infty(\Omega, \mathcal{F}_\tau, P)$ is the Banach algebra of essentially bounded real valued $\mathcal{F}_\tau$-measurable functions.

The Kullback–Leibler distance for the system error will be identified as an essentially bounded $\mathcal{F}_\tau$-measurable function with its class in $L^\infty(\Omega, \mathcal{F}_\tau, P)$. Therefore, an error at a stopping time τ is an element of $L^\infty(\Omega, \mathcal{F}_\tau, P)$. We can distinguish a finite time horizon as a case defined by $\mathcal{F}_t = \mathcal{F}_T \; \forall t \ge T$.

Axiom 3 *For all $\mathcal{K} \in L^\infty(\mathcal{F}_\tau)$, a dynamic error measure $(\varrho_{\tau_1,\tau_2})_{0\le\tau_1\le\tau_2}$ on $(\Omega, \mathcal{F}_\infty, (\mathcal{F}_t)_{t\in\Re^+}, P)$, where $\tau_1 \le \tau_2$ are two stopping times, is a family of maps defined on $L^\infty(\Omega, \mathcal{F}_{\tau_2}, P)$ with values in $L^\infty(\Omega, \mathcal{F}_{\tau_1}, P)$. $(\varrho_{\tau_1,\tau_2})$ is consistent if it satisfies the following properties:*

(1) $\forall \mathcal{K}_1, \mathcal{K}_2, \in L^\infty(\mathcal{F}_\tau)^2$ *such that* $\mathcal{K}_2 \ge \mathcal{K}_1$ *a.s., we have* $\varrho_{\tau_1,\tau_2}(\mathcal{K}_1) \le \varrho_{\tau_1,\tau_2}(\mathcal{K}_2)$.

(2) $\forall \mathcal{K}_1, \mathcal{K}_2, \in L^\infty(\mathcal{F}_\tau)^2$ *such that* $\mathcal{K}_1 + \mathcal{K}_2 \in L^\infty(\mathcal{F}_\tau)^2$, *we have* $\varrho_{\tau_1,\tau_2}(\mathcal{K}_1 + \mathcal{K}_2) \le \varrho_{\tau_1,\tau_2}(\mathcal{K}_1) + \varrho_{\tau_1,\tau_2}(\mathcal{K}_2)$; *for* $\mathcal{K}_1$ *and* $\mathcal{K}_2$ *are comonotonic random variables, i.e., for every* $\omega_1, \omega_2 \in \Omega : (\mathcal{K}_1(\omega_2) - \mathcal{K}_1(\omega_1))(\mathcal{K}_2(\omega_2) - \mathcal{K}_2(\omega_1)) \ge 0$, *we have* $\varrho_{\tau_1,\tau_2}(\mathcal{K}_1 + \mathcal{K}_2) = \varrho_{\tau_1,\tau_2}(\mathcal{K}_1) + \varrho_{\tau_1,\tau_2}(\mathcal{K}_2)$.

(3) $\forall \mathcal{K} \in L^\infty(\mathcal{F}_\tau)$ *and* $\gamma \in \Re_+$ *such that* $\gamma\mathcal{K} \in L^\infty(\mathcal{F}_\tau)$, *we have* $\varrho_{\tau_1,\tau_2}(\gamma\mathcal{K}) \le \gamma\varrho_{\tau_1,\tau_2}(\mathcal{K})$.

(4) $\forall \mathcal{K} \in L^\infty(\mathcal{F}_\tau)$ *and* $a \in L^\infty(\mathcal{F}_\sigma)$ *such that* $\mathcal{K} + a \in L^\infty(\mathcal{F}_\sigma)$, *we have* $\varrho_{\tau_1,\tau_2}(X + a) = \varrho_{\tau_1,\tau_2}(X) + a$.

uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

(5) *For any increasing (resp. decreasing) sequence* $\mathcal{K}_n \in L^\infty(\mathcal{F}_\tau)^n$ *such that* $\mathcal{K} = \lim \mathcal{K}_n$, *the decreasing (resp. increasing) sequence* $\varrho_{\tau_1,\tau_2}(\mathcal{K}_n)$ *has the limit* $\varrho_{\tau_1,\tau_2}(\mathcal{K})$ *and continuous from below (resp. above).*

This axiom is motivated by Axiom 2 in consistency with respect to the dynamic setting for error assessment. Following Bion-Nadal (2008), Chen et al. (2017), and references therein, we can define time invariance as follows:

Definition 3 For all $\mathcal{K} \in L^\infty(\mathcal{F}_\tau)$, a dynamic error measure (ϱ_τ) on $(\Omega, \mathcal{F}_\infty, (\mathcal{F}_t)_{t\in\Re^+}, P)$, where $\tau_1, \tau_2 \in \{0, 1, \ldots, T-1\}$ are consecutive stopping times and s is a constant of time, is said to be time invariant if it satisfies the following properties:

(1) $\forall \mathcal{K}_1, \mathcal{K}_2, \in L^\infty(\mathcal{F}_\tau)^2$: $\varrho_{\tau_1+s}(\mathcal{K}_1) = \varrho_{\tau_1+s}(\mathcal{K}_2) \Rightarrow \varrho_{\tau_1}(\mathcal{K}_1) = \varrho_{\tau_1}(\mathcal{K}_2)$ and $\varrho_{\tau_2+s}(\mathcal{K}_1) = \varrho_{\tau_2+s}(\mathcal{K}_2) \Rightarrow \varrho_{\tau_2}(\mathcal{K}_1) = \varrho_{\tau_2}(\mathcal{K}_2)$.
(2) $\forall \mathcal{K} \in L^\infty(\mathcal{F}_\tau)$: $\varrho_{\tau_1+s}(\mathcal{K}) = \varrho_{\tau_2+s}(\mathcal{K}) \Rightarrow \varrho_{\tau_1}(\mathcal{K}) = \varrho_{\tau_2}(\mathcal{K})$.
(3) $\forall \mathcal{K} \in L^\infty(\mathcal{F}_\tau)$: $\varrho_{\tau_1}(\mathcal{K}) = \varrho_{\tau_1}(-\varrho_{\tau_1+s}(\mathcal{K}))$ and $\varrho_{\tau_2}(\mathcal{K}) = \varrho_{\tau_2}(-\varrho_{\tau_2+s}(\mathcal{K}))$.

Based on the equality and recursive property in Definition 3, we can construct a time invariant error measure by composing one-period (i.e., $s = 1$) measures over time such that for all $t < T-1$ and $\mathcal{K} \in L^\infty(\mathcal{F}_\tau)$ the composed measure $\tilde{\varrho_\tau}$ is (1) $\tilde{\varrho}_{T-1}(\mathcal{K}) = \varrho_{T-1}(\mathcal{K})$ and (2) $\tilde{\varrho}_\tau(\mathcal{K}) = \varrho_\tau(-\tilde{\varrho}_{\tau+1}(\mathcal{K}))$ (see Cheridito and Stadje 2009, for example).

2.3.3 Tail error (TE) and conditional tail error (CTE)

Enlightened by the literature of the extreme value theory, we define the tail error (TE) and its coherent extension, that is, the conditional tail error (CTE). Considering some outcomes generated from a system, an error or distance is measured by the Kullback–Leibler distance $\mathcal{K}$, and there exists any arbitrary value ε, we denote $\mathrm{F}_\mathcal{K}(\varepsilon) = \mathrm{P}(\mathcal{K} \le \varepsilon)$ as the distribution of the corresponding error. We intend to define a statistic based on $\mathrm{F}_\mathcal{K}$ that can measure the severity of the erroneous outcome over a fixed time period. We can directly use the maximum error (e.g., ε), such that $\inf\{\varepsilon \in \Re : \mathrm{F}_\mathcal{K}(\varepsilon) = 1\}$. When the support of $\mathrm{F}_\mathcal{K}$ is not bounded, the maximum error becomes infinity. However, the maximum error does not provide us the probabilistic information of $\mathrm{F}_\mathcal{K}$. The tail error then extends it to the maximum error that will not be exceeded with a given probability.

Definition 4 Given some confidence level $\alpha \in (0, 1)$, the tail error of outcomes from a system at the confidence level α is given by the smallest value ε such that the probability that the error $\mathcal{K}$ exceeds ε is not larger than $1-\alpha$, that is, the tail error is a quantile of the error distribution shown as follows:

$$\begin{aligned} \mathrm{TE}_\alpha(\mathcal{K}) &= \inf\{\varepsilon \in \Re : \mathrm{P}(\mathcal{K} > \varepsilon) \le 1 - \alpha\} \\ &= \inf\{\varepsilon \in \Re : \mathrm{F}_\mathcal{K}(\varepsilon) \ge \alpha\}. \end{aligned} \tag{2}$$

Definition 5 For an error measured by the Kullback–Leibler distance $\mathcal{K}$ with $\mathrm{E}(\mathcal{K}) < \infty$, and a confidence level $\alpha \in (0, 1)$, the conditional tail error (CTE) of system's outcomes at a confidence level α is defined as follows:

$$\mathrm{CTE}_\alpha(\mathcal{K}) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{TE}_\beta\,(\mathcal{K})\,\mathrm{d}\beta, \tag{3}$$

where $\beta \in (0, 1)$.

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

uncorrected proof

We can formally present $\mathrm{TE}_\alpha(\mathcal{K})$ in Eq. (2) as a minimum contrast (discrepancy) parameter as follows:

$$\mathrm{TE}_\alpha(\mathcal{K}) = \arg\min_{\mathrm{TE}_\alpha(\mathcal{K})} \left((1-\alpha) \int_{-\infty}^{\mathrm{TE}_\alpha(\mathcal{K})} |\mathcal{K} - \mathrm{TE}_\alpha(\mathcal{K})| f(\mathcal{K})\, \mathrm{d}\mathcal{K} \right.$$
$$\left. + \alpha \int_{\mathrm{TE}_\alpha(\mathcal{K})}^{\infty} |\mathcal{K} - \mathrm{TE}_\alpha(\mathcal{K})| f(\mathcal{K})\, \mathrm{d}\mathcal{K} \right), \qquad (4)$$

where $\alpha \in (0, 1)$ identifies the location of $\mathcal{K}$ on the distribution of errors. Comparing with TE, CTE averages TE over all levels $\beta \geq \alpha$ and investigates further into the tail of the error distribution. We can see CTE_α is determined only by the distribution of $\mathcal{K}$ and $\mathrm{CTE}_\alpha \geq \mathrm{TE}_\alpha$. We see TE is monotonic, that is, given by Axiom 2-(1), positive homogenous by Axiom 2-(3), translation invariant shown by Axiom 2-(4), and equivalent in probability by Axiom 2-(5). However, the subadditivity property cannot be satisfied in general for TE, and it is thus not a coherent error measure. The conditional tail error satisfies all properties given by Axiom 2 and it is a coherent error measure. In addition, under the dynamic setting, CTE satisfies all properties given by Axiom 3, it is also a dynamic coherent error measure.

In other words, TE measures an error below a given percentage (i.e., α) of all errors. α is then the probability such that an error greater than TE is less than or equal to α, whereas an error less than TE is less or than or equal to $1-\alpha$. CTE measures the average error of the performance in a given percentage (i.e., α) of the worst cases (above TE). Given the expected value of error $\mathcal{K}$ is strictly exceeding TE such that $\mathrm{CTE}_\alpha(\mathcal{K})^+ = \mathrm{E}[\mathcal{K}|\mathcal{K} > \mathrm{TE}_\alpha(\mathcal{K})]$ and the expected value of error $\mathcal{K}$ is weakly exceeding TE where $\mathrm{CTE}_\alpha(\mathcal{K})^- = \mathrm{E}[\mathcal{K}|\mathcal{K} \geq \mathrm{TE}_\alpha(\mathcal{K})]$, $\mathrm{CTE}_\alpha(\mathcal{K})$ can be treated as the weighted average of $\mathrm{CTE}_\alpha(\mathcal{K})^+$ and $\mathrm{TE}_\alpha(\mathcal{K})$ such that:

$$\mathrm{CTE}_\alpha(\mathcal{K}) = \begin{cases} \Phi_\alpha(\mathcal{K})\, \mathrm{TE}_\alpha(\mathcal{K}) + (1 - \Phi_\alpha(\mathcal{K}))\, \mathrm{CTE}_\alpha(\mathcal{K})^+, & \text{if } \mathrm{F}_\mathcal{K}(\mathrm{TE}_\alpha(\mathcal{K})) < 1; \\ \mathrm{TE}_\alpha(\mathcal{K}) & \text{if } \mathrm{F}_\mathcal{K}(\mathrm{TE}_\alpha(\mathcal{K})) = 1, \end{cases} \qquad (5)$$

where

$$\Phi_\alpha(\mathcal{K}) = \frac{\mathrm{F}_\mathcal{K}(\mathrm{TE}_\alpha(\mathcal{K})) - \alpha}{1-\alpha}.$$

We can see that $\mathrm{TE}_\alpha(\mathcal{K})$ and $\mathrm{CTE}_\alpha(\mathcal{K})^+$ are discontinuous functions for the error distributions of $(\mathcal{K})$, whereas $\mathrm{CTE}_\alpha(\mathcal{K})$ is continuous with respect to α; see Sun et al. (2009) and references therein. We then obtain $\mathrm{TE}_\alpha(\mathcal{K}) \leq \mathrm{CTE}_\alpha(\mathcal{K})^- \leq \mathrm{CTE}_\alpha(\mathcal{K})^+ \leq \mathrm{CTE}_\alpha(\mathcal{K})$. It is clear that CTE evaluates the performance quality in a conservative way by focusing on the less accurate outcomes. α for TE and CTE can be used to scale the user's tolerance of critical errors.

2.3.4 Computational methods

The TE_α represents the error that with probability α it will not be exceeded. For continuous distributions, it is an α-quantile such that $\mathrm{TE}_\alpha(\mathcal{K}) = F_\mathcal{K}^{-1}(\alpha)$ where $F_\mathcal{K}(\cdot)$ is the cumulative distribution function of the random variable of error $\mathcal{K}$. For the discrete case, there might not be a defined unique value but a probability mass around the value; we then can compute $\mathrm{TE}_\alpha(\mathcal{K})$ with $\mathrm{TE}_\alpha(\mathcal{K}) = \min\{\mathrm{TE}(\mathcal{K}) : \mathrm{P}(\mathcal{K} \leq \mathrm{TE}(\mathcal{K})) \geq \alpha\}$.

The $\mathrm{CTE}_\alpha(\mathcal{K})$ represents the worst $(1-\alpha)$ part of the error distribution of $\mathcal{K}$ that is above the $\mathrm{TE}_\alpha(\mathcal{K})$ (i.e., conditional on the error above the TE_α). When $\mathrm{TE}_\alpha(\mathcal{K})$ has continuous part of the error distribution, we can derive αth CTE of $(\mathcal{K})$ with $\mathrm{CTE}_\alpha(\mathcal{K}) = \mathrm{E}(\mathcal{K}|\mathcal{K} > \mathrm{TE}_\alpha(\mathcal{K}))$. However, when TE_α falls in a probability mass, we compute $\mathrm{CTE}_\alpha(\mathcal{K})$ following Eq. (5):

uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

$$\mathrm{CTE}_\alpha(\mathcal{K}) = \frac{(\hat{\beta}-\alpha)\,\mathrm{TE}_\alpha(\mathcal{K}) + (1-\hat{\beta})\,\mathrm{E}\,(\mathcal{K}\,|\,\mathcal{K} > \mathrm{TE}_\alpha(\mathcal{K}))}{1-\alpha},$$

where $\hat{\beta} = \max\{\beta : \mathrm{TE}_\alpha(\mathcal{K}) = \mathrm{TE}_\beta(\mathcal{K})\}$.

It is obvious that the α-th quantile of $(\mathcal{K})$ is critical. The classical method for computing is based on the order statistics; see David and Nagaraja (2003) and references therein. Given a set of values $\mathcal{K}_1, \mathcal{K}_1, \ldots, \mathcal{K}_n$, we can define the quantile for any fraction p as follows:

1. Sort the values in order, such that $\mathcal{K}_{(1)} \leq \mathcal{K}_{(2)} \leq \cdots \leq \mathcal{K}_{(n)}$. The values $\mathcal{K}_{(1)}, \ldots, \mathcal{K}_{(n)}$ are called the order statistics of the original sample.
2. Take the order statistics to be the quantile that correspond to the fraction:

$$p_i = \frac{i-1}{n-1},$$

for $i = 1, 2, \ldots, n$.

3. Use the linear interpolation between two consecutive p_i to define the p-th quantile if p lies a fraction f to be:

$$\mathrm{TE}_p(\mathcal{K}) = (1-f)\,\mathrm{TE}_{p_i}(\mathcal{K}) + f\,\mathrm{TE}_{p_{i+1}}(\mathcal{K}).$$

$\mathcal{K}$ in Axiom 1 measures the distance (i.e., error) between $\widetilde{V}(x)$ and $V(x)$. If the error illustrate homogeneity and isotropy, for example, the distribution of errors are characterized by a known probability function, a parametric method then can be applied, see Sun et al. (2009) and references therein.

3 Wavelet-based big data systems

We are now going to evaluate big data systems under the framework we proposed in Sect. 2. We first describe three big data systems that have been used for intelligent decision making with different functions in this section before we conduct a numerical investigation. These three big data systems have been built with wavelet algorithms that were applied in big data analytics; see Sun and Meinl (2012) and references therein. The first big data system is named SOWDA (smoothness-oriented wavelet denoising algorithm), which optimizes a decision function with two control variables as proposed by Chen et al. (2015). It is originally used to preserve the smoothness (low-frequency information) when processing big data. The second is GOWDA (generalized optimal wavelet decomposition algorithm) introduced by Sun et al. (2015), which is an efficient system for data processing with multiple criteria (i.e., six equally weighted control variables). Because GOWDA has a more sophisticated decision function, an equal weight framework significantly improves computational efficiency. It is particularly suitable for decomposing multiple-frequency information with heterogeneity. The third one is called WRASA (wavelet recurrently adaptive separation algorithm), an extension of GOWDA as suggested by Chen et al. (2015), which determines the decision function of multiple criteria with convex optimization to optimally weight the criteria. Chen and Sun (2018) illustrate how to incorporate these systems with a reinforcement learning framework. All three big data systems have open access to adopt components of wavelet analysis.

As we have discussed in Sect. 2, a complete BDS contains two technology stacks: data engineering and data science. The former intends to design, develop, and integrate data from various resources and run ETL on top of big datasets and the latter applies data mining,

uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

machine learning, and statistics approaches to discovering knowledge and developing intelligence (such as decision making). In this section, we briefly introduce the three BDS focusing on their engineering and learning.

3.1 Data engineering with wavelet

Big data has several unique characteristics (see Sect. 2.1) and requires special processing before mining it. When dealing with heterogeneity, multi-resolution transformation of the input data enables us to analyze the information simultaneously on different dimensions. We are going to decompose the input data X as follows:

$$X_t = S_t + N_t, \tag{6}$$

where S_t is the trend and we can estimate it with $\tilde{S}_t$ and N_t is the additive random features sampled at time t. Let $\mathcal{W}(\cdot; \omega, \zeta)$ denote the wavelet transform operators with specific wavelet function ω for ζ level of decomposition and $\mathcal{W}^{-1}(\cdot)$ is its corresponding inverse transform. Let $\mathcal{D}(\cdot; \gamma)$ denote the denoising operator with thresholding rule γ. Therefore, wavelet data engineering is then to extract $\tilde{S}_t$ from X_t as an optimal approximation of S_t after removing the randomness. We summarize the processing as follows:

$$Y = \mathcal{W}(X; \omega, \zeta),$$
$$Z = \mathcal{D}(Y; \gamma),$$
$$\tilde{S} = \mathcal{W}^{-1}(Z).$$

The complete procedure shall implement the operators $\mathcal{W}(\cdot)$ and $\mathcal{D}(\cdot)$ after careful selection of ω, ζ, and γ; see Sun et al. (2015).

Wavelets are bases of $L^2(\mathbb{R})$ and enable a localized time-frequency analysis with wavelet functions that usually have either compact support or decay exponentially fast to zero. Wavelet function (or the mother wavelet) $\psi(t)$ is defined as

$$\int_{-\infty}^{\infty} \psi(t)dt = 0.$$

Given $\alpha, \beta \in \mathbb{R}, \alpha \neq 0$, with compressing and expanding $\psi(t)$, we obtain a successive wavelet $\psi_{\alpha,\beta}(t)$, such that,

$$\psi_{\alpha,\beta}(t) = \frac{1}{\sqrt{|\alpha|}} \psi\left(\frac{t-\beta}{\alpha}\right), \tag{7}$$

where we call α the scale or frequency factor and β the time factor. Given $\bar{\psi}(t)$ as the complex conjugate function of $\psi(t)$, a successive wavelet transform of the time series or finite signals $f(t) \in L^2(\mathbb{R})$ can be defined as

$$\mathcal{W}_{\psi} f(\alpha, \beta) = \frac{1}{\sqrt{|\alpha|}} \int_{\mathbb{R}} f(t)\bar{\psi}\left(\frac{t-\beta}{\alpha}\right) dt. \tag{8}$$

Equation (8) shows that the wavelet transform is the decomposition of $f(t)$ under a different resolution level (scale). In other words, the idea of the wavelet transform $\mathcal{W}$ is to translate and dilate a single function (the mother wavelet) $\psi(t)$. The degree of compression is determined by the scaling parameter α, and the time location of the wavelet is set by the translation parameter β. Here, $|a| < 1$ leads to compression and thus higher frequencies. The opposite (lower frequencies) is true for $|a| > 1$, leading to time widths adapted to their frequencies, see Chen et al. (2017) and references therein.

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

uncorrected proof

The continuous wavelet transform (CWT) is defined as the integral over all time of the data multiplied by scaled (stretching or compressing), shifted (moving forward or backward) version of wavelet function. Because most big data applications have a finite length of time, therefore, only a finite range of scales and shifts are practically meaningful. Therefore, we consider the discrete wavelet transform (DWT), which is an orthogonal transform of a vector (discrete big data) X of length N (which must be a multiple of 2^J) into J wavelet coefficient vectors $W_j \in \mathbb{R}^{N/2^j}$, $1 \leq j \leq J$ and one scaling coefficient vector $V_J \in \mathbb{R}^{N/2^J}$. The maximal overlap discrete wavelet transform (MODWT) has also been applied to reduce the sensitivity of DWT when choosing an initial value. The MODWT follows the same pyramid algorithm as the DWT by using the rescaled filters; see Sun et al. (2015) for details. All three BDSs can apply the DWT and MODWT. We focus on the wavelet transform on MODWT in this study.

Sun et al. (2015) pointed out that wavelet transform can lead to different outcomes when choosing different input variables. Since wavelets are oscillating functions and transform the signal orthogonally in the Hilbert space, there are many wavelet functions that satisfy the requirement for the transform. Chen et al. (2015) show that different wavelets are used for different reasons. For example, the Haar wavelet has very small support, whereas wavelets of higher orders such as Daubechies DB(4) and least asymmetric (LA) have bigger support. Bigger support can ensure a smoother shape of S_j, for $1 \leq j \leq J$ with each wavelet and that the scaling coefficient carries more information because of the increased filter width. Sun et al. (2015) highlighted three factors that influence the quality of wavelet transform: wavelet function (or mother wavelet), number of maximal iterations (or level of decomposition), and the thresholding rule. However, there is no straightforward method to determine these three factors simultaneously. Therefore, different BDS applies different decision functions when conducting machine learning.

3.2 Machine learning

In order to perform the wavelet transform to optimize the analysis, that is, to see how close $\tilde{S}_t$ is toward S_t, three big data systems have been applied with different decision processing. First, let us name a separation factor θ, which is the combination of wavelet function, number of maximal iterations, and thresholding rule, as well as a factor space $\Theta \subseteq \mathbb{R}^p$, $p \geq 0$, then $\theta \in \Theta$. Different θ will lead to different wavelet performances. We then consider a random variable $\mathcal{K}$ and $\mathcal{K} = S - \tilde{S}$. Here, $\mathcal{K}$ is in fact the approximation error and determined by a real value approximating function $m(\mathcal{K}, \theta) \in \mathcal{R}$ and $E(m(\mathcal{K}, \theta)) = 0$.

Definition 6 Let $\mathcal{K}_t \in \mathbb{R}^d$ be i.i.d. random vectors, and given that $\theta \in \mathbb{R}^p$ is uniquely determined by $E(m(\mathcal{K}, \theta)) = 0$, where $m(\mathcal{K}, \theta)$ is called the approximating function and takes values in $\mathbb{R}^{p+q}$ for $q \geq 0$. Algorithm determines $\tilde{\theta}$ such that $\tilde{\theta} = \arg\max_\theta \mathcal{R}(\theta)$, where

$$\mathcal{R}(\theta) = \min \left\{ \sum_{i=1}^{n} \omega_i \, m_i(\mathcal{K}, \theta) \right\}, \quad s.t. \, \omega_i \geq 0; \; \sum_{i=1}^{n} \omega_i = 1, \tag{9}$$

where ω_i is the weight for the approximating function.

If we use only one type of approximation—that is, a single criterion—then the weight is one. If we need to apply multivariate criteria for approximation, then we have to determine the associated weight for each approximating function.

uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

3.2.1 Decision criteria

Six model selection criteria (i.e., $m_1, \ldots, m_6$) have been suggested by Sun et al. (2015) to evaluate $m(\cdot)$. Given $\mathcal{K}_t = S_t - \tilde{S}_t$, for all $t = 1, 2, \ldots, T$ where T is the data length and p the number of parameters estimated, the first three criteria are shown as follows:

$$m_1 = \sqrt{\frac{\sum_{t=1}^{T} \mathcal{K}_t^2}{T}}, \tag{10}$$

$$m_2 = \ln\left(\frac{\sum_{t=1}^{T} \mathcal{K}_t^2}{T}\right) + \frac{2p}{T}, \tag{11}$$

and

$$m_3 = \ln\left(\frac{\sum_{t=1}^{T} \mathcal{K}_t^2}{T}\right) + \frac{p \ln T}{T}. \tag{12}$$

Another three criteria are based on specific indicating functions. They are given as follows:

$$m_4 = \frac{1}{T} \sum_{t=1}^{T} 1 \times \mathbb{1}_{\mathcal{C}(\mathcal{K})=1}, \tag{13}$$

where

$$\mathcal{C}(\mathcal{K}) = \begin{cases} 1, & if \quad \frac{\max |\mathcal{K}_t - \frac{1}{T}\sum_1^T \mathcal{K}_t|}{\sqrt{\frac{1}{T-1}\sum_1^T (\mathcal{K}_t - \frac{1}{T}\sum_1^T \mathcal{K}_t)^2}} > z_\alpha \\ 0, & otherwise \end{cases}$$

and z is the α-th quantile of a probability distribution (e.g. $\alpha = 0.05$) for $\mathcal{K}$.

$$m_5 = \frac{1}{T} \sum_{t=1}^{T} 1 \times \left(\mathbb{1}_{\mathcal{D}(\mathcal{K})=1} + \mathbb{1}_{\mathcal{D}^*(\mathcal{K})=1}\right) \tag{14}$$

where

$$\mathcal{D}(\mathcal{K}) = \begin{cases} 1, & if \quad \mathcal{K}_t \leq \Lambda \\ 0, & if \quad otherwise \end{cases}$$

when Λ is the local maxima and

$$\mathcal{D}^*(\mathcal{K}) = \begin{cases} 1, & if \quad \mathcal{K}_t \geq \lambda \\ 0, & if \quad otherwise \end{cases}$$

when λ is the local minima.

$$m_6 = \frac{1}{T} \sum_{t=1}^{T} 1 \times \mathbb{1}_{\mathcal{C}(\mathcal{K})=1}, \tag{15}$$

where

$$\mathcal{C}(\mathcal{K}) = \begin{cases} 1, & if \quad (S_{t+1} - S_t)(\tilde{S}_{t+1} - S_t) \leq 0 \\ 0, & if \quad otherwise. \end{cases}$$

Summaried by Sun et al. (2015) is m_4 to indicate the global extrema and m_5 for the local extrema. Both of them have the ability to detect boundary problems—that is, an inefficient

uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

approximation at the beginning and the end of the signal. In addition, m_6 detects the severity of $\mathcal{K}$ focusing on directional consistence.

For the three big data systems we compare in our study, SOWDA uses m_1 and m_2 and GOWDA and WRASA use all six criteria when formulating their decision making process.

3.2.2 Decision making

With the approximation error vector $\mathbf{m} = [m_1, m_2, \ldots, m_6]$, the next step is to determine the weights of these factors, that is, $\boldsymbol{\omega} \in \mathbb{R}^6$. Because the approximation error $\mathbf{m}$ is random, for any given combination of the wavelet, level of decomposition, and thresholding function, the weighted error $\mathbf{m}^T\boldsymbol{\omega}$ can be expressed as a random variable with mean $\mathbb{E}[\mathbf{m}^T\boldsymbol{\omega}]$ and variance $\mathbb{V}(\mathbf{m}^T\boldsymbol{\omega})$ where $\mathbb{E}[\mathbf{m}^T\boldsymbol{\omega}] = \bar{\mathbf{m}}^T\boldsymbol{\omega}$ and $\mathbb{V}(\mathbf{m}^T\boldsymbol{\omega}) = \mathbb{E}[\mathbf{m}^T\boldsymbol{\omega} - \mathbb{E}[\mathbf{m}^T\boldsymbol{\omega}]]^2 = \boldsymbol{\omega}^T\Sigma\,\boldsymbol{\omega}$. In our work, we apply convex optimization suggested by Chen et al. (2017) to determine the weight $\boldsymbol{\omega}$ for minimizing a linear combination of the expected value and variance of the approximation error, that is,

$$\begin{aligned} \underset{\boldsymbol{\omega}}{\text{minimize}} \quad & \mathbf{m}^T\boldsymbol{\omega} + \gamma\, \boldsymbol{\omega}^T \Sigma\, \boldsymbol{\omega} \\ \text{s. t.} \quad & \mathbf{1}^T\boldsymbol{\omega} = 1, \\ & \boldsymbol{\omega} \succeq 0, \end{aligned}$$

where γ is a sensitivity measure that trades off between small expected approximation error and small error variance because of computational efficiency. In this article, we allow $\gamma > 0$ which ensures a sufficiently large decrease in error variance, when the expected approximation error is trivial with respect to computational tolerance.

Figure 2 illustrates the workflow for the three big data systems (i.e., SOWDA, GOWDA and WRASA) that we investigated in this article. As we can see, the major difference of these three big data systems is in the decision-making process. WRASA minimizes the resulting difference sequence between $\tilde{S}_t$ and S_t based on choosing an optimal combination of approximating functions $m(\cdot)$. Therefore, WRASA determines the approximating function $m(\cdot)$ and its corresponding weight with convex optimization. SOWDA and GOWDA simply take the predetermined value of ω_i (i.e., at equal weight in our study) rather than optimizing them. There has to be a trade-off between accuracy and computational efficiency if we want to keep the latency as low as possible.

4 Empirical study

4.1 The data

In our empirical study, we use electricity power consumption in France from 01 January 2009 to 31 December 2014. The consumption data are taken from real-time measurements of production units. The National Control Center (Centre National d'Exploitation du Système - CNES) constantly matches electricity generation with consumers' power demands, covering total power consumed in France nationwide (except Corsica). The input data for our empirical study are provided by RTE (Ré seau de Transport d'Èlectricitè—Electricity Transmission Network), which is the electricity transmission system operator of France. RTE is in charge of operating, maintaining and developing high-voltage transmission systems in France, which is also the Europe's largest one at approximately 100,000 kilometers in length. The data

Uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

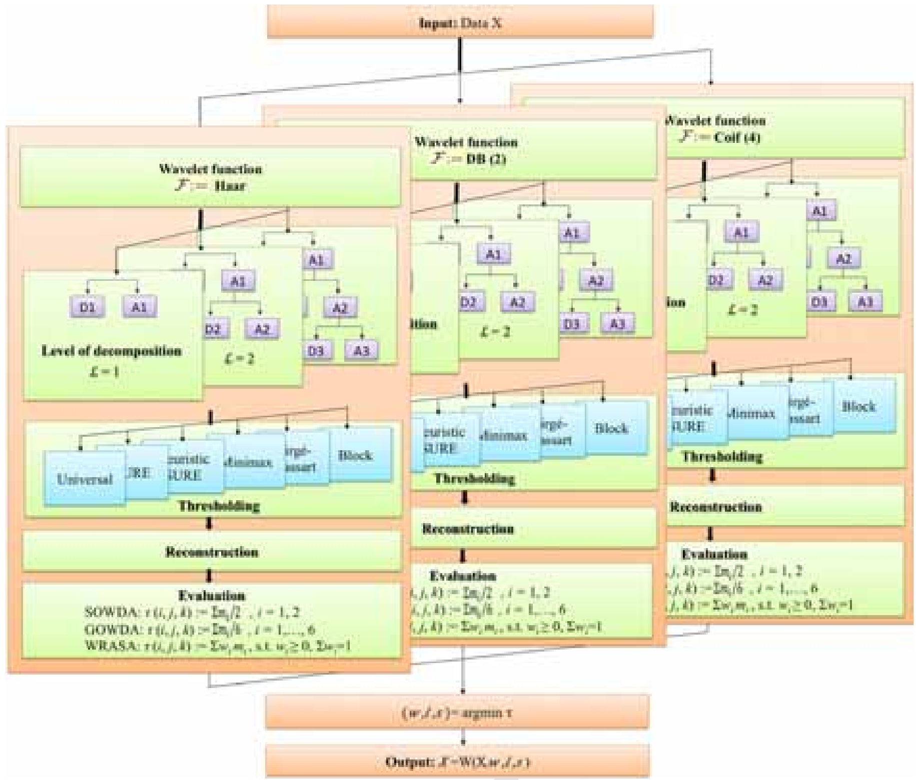


Fig. 2 Workflow of three big data systems (SOWDA, GOWDA, and WRASA) that have different decision processing paths at the evaluation stage

supplied are calculated on the basis of load curves at 10-min frequency. RTE's regional units validate the data by adding missing values and by correcting erroneous data then provide the information data at 30-min frequency. These data are completed by the estimated values for consumed power from electricity generation on the distribution networks and private industrial networks.[4]

4.2 The method

In our empirical study, we apply three big data systems (i.e., SOWDA, GOWDA, and WRASA) to complete the whole decision making process, that is, process the original power demand data and conduct a predictive analytics. We evaluate the quality of decision-making process (forecasting accuracy) with the TE and CTE discussed in Sect. 2.3.

Following Sun et al. (2015) and Chen et al. (2017), we chose Haar, Daubechies (DB), Symlet (LA), and Coiflets (Coif) as wavelet functions. We apply the MODWT with pyramid algorithm in our empirical study. Several threshold rules have been considered as well. The components served for the three systems in our study are listed as follows:

- $\omega \in$ { Haar, DB(2), DB(4), DB(8), LA(2), LA(4), LA(8), Coif(4), Coif(6), Coif(8)};
- $\zeta \in \{i : i = 1, 2, 3\}$;

[4] See www.rte-france.com.

- $\gamma \in$ {Heuristic SURE, Minimax, SURE, Universal}.

The subsample series that are used for the in-sample study are randomly selected by a moving window with length T. Replacement is allowed in the sampling. Letting T_F denote the length of the forecasting series, we perform the 1 day ahead out-of-sample forecasting ($1 \leq T \leq T + T_F \leq N$). In our analysis, the total sample length is $T = 17{,}520$ for a normal year and $T = 17{,}568$ for a leap year. The subsample length (i.e., the window length) of $T = 672$ (2 weeks) was chosen for the in-sample simulation and $T_F = 48$ (1 day) for the out-of-sample forecasting.

We compare the system's performance by using the root mean squared error (RMSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE) as the goodness-of-fit measures:

$$\text{RMSE} := \sqrt{\frac{\sum_{t=1}^{T}\left(S_t - \tilde{S}_t\right)^2}{T}},$$

$$\text{MAE} := \frac{\sum_{t=1}^{T}|S_t - \tilde{S}_t|}{T}, \text{ and}$$

$$\text{MAPE} := \frac{100}{T}\sum_{t=1}^{T}\left|\frac{S_t - \tilde{S}_t}{S_t}\right|,$$

where S_t denotes the actual value at time t, and $\tilde{S}_t$ is its corresponding predicted value generated by the big data system.

Baucells and Borgonovo (2013) and Sun et al. (2007) suggested the Kolmogorov–Smirnov (KS) distance, the Cramér–von Mises (CVM) distance, and the Kuiper (K) distance as the metrics for evaluation. We use them here to compare the quality performance of the big data systems in our study.

Let $F_n(S)$ denote the actual sample distribution of S, and $F(\tilde{S})$ is the distribution function of the approximation or forecasts (i.e., the output of BDS), Kolmogorov–Smirnov (KS) distance, Cramér von Mises (CVM), and Kuiper distance (K), which are defined as follows:

$$\text{KS} := \sup_{x\in\mathbb{R}}\left|F_n(S) - F(\tilde{S})\right|,$$

$$\text{CVM} := \int_{-\infty}^{\infty}\left(F_n(S) - F(\tilde{S})\right)^2 \mathrm{d}F(\tilde{S}), \text{ and}$$

$$\text{K} := \sup_{x\in\mathbb{R}}\left(F_n(S) - F(\tilde{S})\right) + \sup_{x\in\mathbb{R}}\left(F(\tilde{S}) - F_n(S)\right).$$

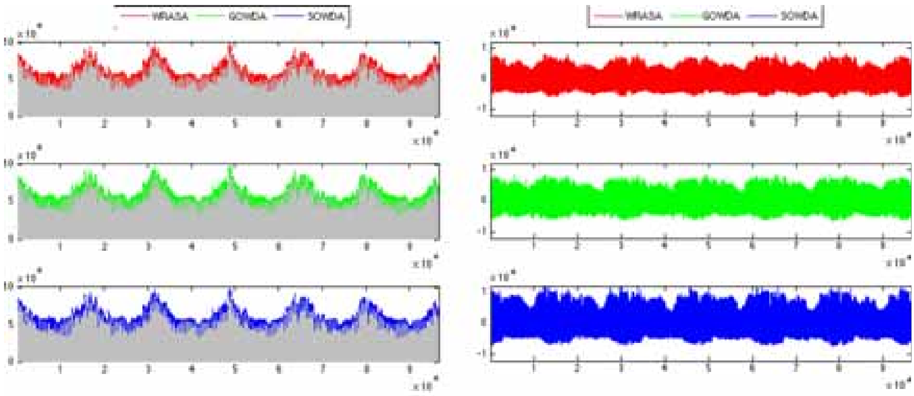
We conclude that the smaller these distances are, the better the performance of the system. The KS distance focuses on deviations near the median of the distribution. Thus, it tends to be more sensitive near the center of the error distribution than at the tails, whereas the CVM distance measures the sensitivity of dispersion between the output and its trend with respect to the change of the data filter of BDS. The Kuiper distance considers the extreme errors.

4.3 Results

We show the forecasting performance of three big data systems (i.e., SOWDA, GOWDA, and WRASA) in Fig. 3. Panel (a) illustrates the predictive power demand generated from the three big data systems when compared with the actual power demand. Panel (b) illustrates

uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small



(a) Forecasting with three BDSs (b) Forecasting error of three BDSs

Fig. 3 Comparison for forecasting performance of the three big data systems (i.e., SOWDA, GOWDA, and WRASA). **a** Illustrates the predictive electricity power demand generated from the three systems compared with the actual electricity power demand. **b** Illustrates the forecasting errors compared with the actual electricity power demand

the forecasting errors versus the actual power demand. It is not easy to identify which one performs better than the others. Therefore, we report different error evaluation results.

Table 1 reports the in-sample (training) and out-of-sample (forecasting) performances of SOWDA, GOWDA, and WRASA measured by MAPE, MAE, and RMSE as the goodness-of-fit criteria. For the in-sample modeling performances, we see that the values of MAPE, MAE, and RMSE on average are all reduced after applying WRASA, compared with the results of SOWDA and GOWDA. Compared to SOWDA, WRASA reduces the error on average by 27.75, 27.87, and 24.17% for the MAPE, MAE, and RMSE criteria, respectively. Compared to GOWDA, WRASA reduces the error on average by 24.55, 24.56, and 21.47% for the MAPE, MAE, and RMSE, respectively.

We also focus on the results of the out-of-sample forecasting performances of SOWDA, GOWDA, and WRASA and evaluate them by using MAPE, MAE, and RMSE as the quality criteria. For the forecasting results, WRASA increases the accuracy on average by 28.06, 28.13, and 28.95% compared with SOWDA, and 10.49, 10.12, and 10.22% compared with GOWDA, based on the MAPE, MAE, and RMSE criteria, respectively.

Table 2 reports the in-sample and out-of-sample performance of SOWDA, GOWDA, and WRASA when we take the Cramér von Mises (CVM), Kolmogorov–Smirnov (KS), and Kuiper statistics into consideration as the goodness-of-fit criteria. In comparison with the results of SOWDA, WRASA improves the performance by 56.77, 38.26, and 32.94% measured by the CVM, KS, and Kuiper criteria, respectively. In addition, WRASA reduces the error by 49.84, 19.79, and 24.54% compared with GOWDA when evaluating by the CVM, KS, and Kuiper distance, respectively. From the CVM, KS, and Kuiper criteria, WRASA achieves better forecasting performance than SOWDA by 54.99, 34.58, and 30.80%, respectively. Under the same measurement, WRASA outperforms GOWDA by 31.41, 10.88, and 6.84%, respectively.

We next evaluate the performances of three big data systems with the methods we have discussed in Sect. 2.3—that is, TE and CTE. With these tools, we can identify the performance quality with respect to their cumulative absolute errors. We choose the significance level α at 50%, 25%, 10%, 5%, and 1%. When $\alpha = 50\%$, TE is then the median of the absolute

Uncorrected Proof

Table 1 Comparison of the in-sample (modeling) and out-of-sample (forecasting) performances measured by the mean absolute percentage error (MAPE), the mean absolute error (MAE), and the root mean squared error (RMSE) for three big data systems (i.e., SOWDA, GOWDA, and WRASA) with the electricity power consumption data in France at 30-min frequency from 01 January, 2009 to 31 December, 2014

Year	SOWDA			GOWDA			WRASA		
	MAPE[a]	MAE[b]	RMSE[b]	MAPE[a]	MAE[b]	RMSE[b]	MAPE[a]	MAE[b]	RMSE[b]
In-sample									
2009	3.0185	1.5870	1.9814	2.8154	1.4784	1.8749	2.3259	1.2258	1.5974
2010	2.7178	1.5053	1.8968	2.6288	1.4508	1.8437	1.7450	0.9568	1.2757
2011	2.8850	1.5055	1.9051	2.7805	1.4483	1.8479	1.9873	1.0327	1.3910
2012	2.9668	1.5660	1.9620	2.8594	1.5062	1.9074	2.1041	1.1126	1.4665
2013	2.9423	1.5721	1.9562	2.8441	1.5206	1.9067	2.2082	1.1757	1.5353
2014	3.1117	1.5779	1.9700	2.9642	1.5012	1.8895	2.3752	1.2146	1.5847
Mean	2.9404	1.5523	1.9452	2.8154	1.4842	1.8784	2.1243	1.1197	1.4751
SD	0.1332	0.0370	0.0355	0.1104	0.0301	0.0280	0.2340	0.1073	0.1245
Out-of-sample									
2009	3.9534	2.0636	2.7601	3.2247	1.6737	2.2063	2.7910	1.4596	1.9417
2010	3.6654	2.0143	2.6868	2.8704	1.5763	2.0847	2.5661	1.4100	1.8510
2011	3.7952	1.9664	2.6387	3.0108	1.5521	2.0652	2.7307	1.4109	1.8618
2012	3.9094	2.0533	2.7274	3.1387	1.6432	2.1584	2.8005	1.4724	1.9334
2013	3.9231	2.0869	2.7522	3.1521	1.6686	2.1840	2.8560	1.5162	1.9886
2014	4.0341	2.0408	2.7045	3.3145	1.6618	2.1762	3.0037	1.5172	1.9829
Mean	3.8801	2.0375	2.7116	3.1185	1.6293	2.1458	2.7913	1.4644	1.9266
SD	0.1305	0.0424	0.0452	0.1577	0.0520	0.0573	0.1441	0.0477	0.0587

[a] and [b] denote $\times 10^{-2}$ and $\times 10^{3}$ respectively

error. Table 3 shows the comparison of SOWDA, GOWDA, and WRASA for their training performance measured by TE and CTE, and Table 4 shows the comparison of forecasting performance. 4

As we have mentioned, TE is exactly the α-th quantile of the error distribution. We could also say that, the smaller TE is, the better the system's performance will be. CTE assesses the average error between TE at α confidence and the errors exceeding it. Obviously, the smaller CTE is, the better the system's performance is as well.

We find that the values of TE and CTE for WRASA are the smallest compared with that of SOWDA and GOWDA for all αs in Tables 3 and 4. For example, in Table 3 when we compare the training performance with TE, for $\alpha = 5\%$ WRASA reduces the tail error on average by 21.43 and 16.58% compared with the results of SOWDA and GOWDA, respectively. When $\alpha = 1\%$, the reduction of WRASA then turns to be 15.78 and 15.12%, respectively. For CTE, when $\alpha = 5\%$ WRASA reduces the conditional tail error on average by 17.63 and 15.47% in comparison with SOWDA and GOWDA, respectively and when $\alpha = 1\%$, WRASA reduces the conditional tail error by 12.62 and 11.73%, respectively.

When we compare the forecasting performance in Table 4, for $\alpha = 5\%$ WRASA reduces TE by 27.49 and 10.91% comparing with SOWDA and GOWDA and CTE by 43.36 and 8.91%, respectively. For $\alpha = 1\%$ WRASA reduces TE by 33.10 and 7.6% versus SOWDA and GOWDA and CTE 49.15 and 6.13%, respectively.

uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

Table 2 Comparison of the in-sample (modeling) and out-of-sample (forecasting) performances measured by the Kolmogorov–Smirnov (KS) distance, the Cramér–von Mises (CVM) distance, and the Kuiper (K) distance for three big data systems (i.e., SOWDA, GOWDA, and WRASA) with the electricity power consumption data in France at 30-min frequency from 01 January, 2009 to 31 December, 2014

Year	SOWDA			GOWDA			WRASA		
	CVM[a]	KS[b]	Kuiper[a]	CVM[a]	KS[b]	Kuiper[a]	CVM[a]	KS[b]	Kuiper[a]
In-sample									
2009	25.6560	19.7500	2.8369	12.8150	11.0740	1.9408	12.2280	11.9870	1.9405
2010	4.8961	7.8772	1.2786	6.7468	8.4480	1.4156	1.5314	4.6236	0.7535
2011	15.1470	14.3840	2.2547	13.4820	11.017	2.0207	5.2405	8.3909	1.4327
2012	7.5085	10.4170	1.7761	8.9619	9.1080	1.8159	2.1589	5.8633	0.9962
2013	7.6970	10.1600	1.6553	9.0182	8.8475	1.5754	4.3538	6.9639	1.3243
2014	10.9350	11.5300	1.9408	10.8870	8.5621	1.6668	5.5425	7.9342	1.4270
Mean	11.9733	12.3530	1.9571	10.3185	9.5094	1.7392	5.1759	7.6272	1.3124
SD	7.5720	4.1978	0.5380	2.5633	1.2119	0.2292	3.8192	2.5400	0.4086
Out-of-sample									
2009	5.1905	8.0846	1.3868	4.7813	6.6542	1.1505	3.1291	6.1567	1.2313
2010	3.6509	6.0945	1.2065	2.7960	5.4104	1.0759	1.6588	5.0373	0.9826
2011	8.1781	9.2040	1.7786	5.0804	5.9080	1.1754	3.1970	5.2861	1.0323
2012	7.5418	8.7426	1.5253	4.0844	6.3864	1.1719	2.9793	5.2703	1.0355
2013	3.8286	7.8980	1.4117	3.1775	5.8458	1.0199	2.1897	5.2239	0.9764
2014	8.4888	9.7637	1.8719	4.2839	6.3433	1.2251	3.4463	5.5970	1.0945
Mean	6.1465	8.2979	1.5301	4.0339	6.0914	1.1365	2.7667	5.4286	1.0588
SD	2.1943	1.2828	0.2521	0.8927	0.4528	0.0749	0.6905	0.3997	0.0947

[a] and [b] denote $\times 10^{-2}$ and $\times 10^{-3}$, respectively

As we have pointed out that, when $\alpha = 0.5$, $\text{TE}_{0.5}$ becomes the median of absolute error. Comparing with MAE, we can identify the skewness of the underlying error distribution. When the mean is larger than the median, we call it positively skewed (i.e., the mass of the error is concentrated on the left approaching to zero). The better the performance of the BDS, the more positive skewness of the mass of error approaching to zero. On the other hand, given a skewness, we prefer the light tail (i.e., small tail or conditional tail error). Our results in Tables 3 and 4 illustrate that the error distributions of three systems are all positively skewed to zero.

In addition, from Tables 3 and 4, we can sketch the error distribution above the median. For any given two confidences α_1 and α_2, if $\alpha_1 > \alpha_2$, we can ensure that $\text{TE}_{(1-\alpha_1)}(\mathcal{K}) \leq \text{TE}_{(1-\alpha_2)}(\mathcal{K})$ and $\text{CTE}_{(1-\alpha_1)}(\mathcal{K}) \leq \text{CTE}_{(1-\alpha_2)}(\mathcal{K})$, see Sect. 2.3.3. This is a robust feature for TE and CTE. Furthermore, TE and CTE show their efficiency. When we compare two performance ($\mathcal{K}_1$ and $\mathcal{K}_2$), there must exist a α such that $\text{TE}_{(1-\alpha)}(\mathcal{K}_1) \neq \text{TE}_{(1-\alpha)}(\mathcal{K}_2)$ and $\text{CTE}_{(1-\alpha)}(\mathcal{K}_1) \neq \text{CTE}_{(1-\alpha)}(\mathcal{K}_2)$ if the underlying performance are not identical. With these features, we can compare the system performance by choosing a proper α that reflects the decision maker's utility. For example, when we choose $\text{TE}_{0.9}$, we can conclude that WRASA performs better than GWODA and SOWDA in both in-sample training and out-of-sample forecasting because the corresponding tail and conditional tail errors of WRASA are smallest. Similarly, when we apply TE and CTE for all α values, we conclude the same.

Uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

Table 3 Quality evaluation with TE and CTE at different confidence levels for in-sample (training) performances of three big data systems (i.e., SOWDA, GOWDA, and WRASA) with the electricity power consumption data in France at 30-min frequency from 01 January, 2009 to 31 December, 2014

Year	SOWDA[a]					GOWDA[a]					WRASA[a]				
	$TE_{0.5}$	$TE_{0.75}$	$TE_{0.9}$	$TE_{0.95}$	$TE_{0.99}$	$TE_{0.5}$	$TE_{0.75}$	$TE_{0.9}$	$TE_{0.95}$	$TE_{0.99}$	$TE_{0.5}$	$TE_{0.75}$	$TE_{0.9}$	$TE_{0.95}$	$TE_{0.99}$
2009	1.3456	2.2091	3.2774	3.9666	5.2391	1.2166	2.1276	3.0656	3.6961	5.1308	0.9584	1.7209	2.6513	3.3097	4.5282
2010	1.2404	2.0868	3.1486	3.8936	5.0365	1.1708	2.0962	3.0512	3.6441	4.9777	0.7438	1.2865	2.0687	2.6557	4.0111
2011	1.2192	2.1078	3.1905	3.8674	5.0959	1.1452	2.0869	3.0512	3.6593	5.0578	0.7653	1.3920	2.3264	2.9903	4.1583
2012	1.3033	2.2083	3.2398	3.9454	5.1505	1.2267	2.1823	3.1172	3.7574	5.1653	0.8554	1.5282	2.4306	3.0448	4.3268
2013	1.3124	2.2098	3.2662	3.9139	4.9845	1.2516	2.1944	3.1348	3.7037	5.0133	0.8883	1.6811	2.5566	3.1990	4.3516
2014	1.3189	2.2223	3.2487	3.9064	5.1361	1.2286	2.1630	3.0793	3.6665	5.0578	0.9075	1.7580	2.6522	3.2596	4.4311
Mean	1.2900	2.1740	3.2285	3.9156	5.1071	1.2066	2.1417	3.0832	3.6878	5.0671	0.8531	1.5611	2.4476	3.0765	4.3012
Year	$CTE_{0.5}$	$CTE_{0.75}$	$CTE_{0.9}$	$CTE_{0.95}$	$CTE_{0.99}$	$CTE_{0.5}$	$CTE_{0.75}$	$CTE_{0.9}$	$CTE_{0.95}$	$CTE_{0.99}$	$CTE_{0.5}$	$CTE_{0.75}$	$CTE_{0.9}$	$CTE_{0.95}$	$CTE_{0.99}$
2009	2.4989	3.2512	4.1641	4.7340	5.7030	2.3691	3.1023	3.9454	4.5507	5.5692	1.9842	2.6805	3.5075	4.0747	5.2337
2010	2.3859	3.1351	4.0502	4.5939	5.4996	2.3317	3.0676	3.8806	4.4553	5.4511	1.5606	2.1265	2.8810	3.4189	4.5978
2011	2.3974	3.1600	4.059	4.6278	5.6314	2.3300	3.0778	3.9178	4.5216	5.5630	1.6971	2.3523	3.1829	3.7379	4.8103
2012	2.4786	3.2347	4.1217	4.6755	5.6412	2.4151	3.1643	3.9864	4.5906	5.6334	1.8059	2.4698	3.2699	3.8343	4.9502
2013	2.4742	3.2173	4.0751	4.5772	5.4823	2.4221	3.1464	3.9378	4.5082	5.4949	1.9124	2.5967	3.3680	3.8898	4.7809
2014	2.4925	3.2418	4.1051	4.6440	5.6451	2.3986	3.1151	3.9236	4.5149	5.5543	1.9751	2.6820	3.4611	3.9871	4.9906
Mean	2.4546	3.2067	4.0959	4.6421	5.6004	2.3778	3.1122	3.9320	4.5236	5.5443	1.8225	2.4846	3.2784	3.8238	4.8939

[a] denotes $\times 10^3$

uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

Table 4 Quality evaluation with TE and CTE at different confidence levels for out-of-sample (forecasting) performances of three big data systems (i.e., SOWDA, GOWDA, and WRASA) with the electricity power consumption data in France at 30-min frequency from 01 January, 2009 to 31 December, 2014

Year	SOWDA[a]					GOWDA[a]					WRASA[a]				
	$TE_{0.5}$	$TE_{0.75}$	$TE_{0.9}$	$TE_{0.95}$	$TE_{0.99}$	$TE_{0.5}$	$TE_{0.75}$	$TE_{0.9}$	$TE_{0.95}$	$TE_{0.99}$	$TE_{0.5}$	$TE_{0.75}$	$TE_{0.9}$	$TE_{0.95}$	$TE_{0.99}$
2009	1.5508	2.9490	4.3347	5.7975	8.5595	1.2841	2.2885	3.8948	4.7988	6.1450	1.1165	1.9759	3.2585	4.2130	5.6639
2010	1.5105	2.8968	4.2803	5.6249	8.3270	1.1984	2.1694	3.6167	4.5703	5.8437	1.0944	1.9494	3.1243	3.9322	5.3448
2011	1.4615	2.8115	4.1919	5.6392	8.1671	1.1669	2.1635	3.5597	4.4997	5.8968	1.0863	1.9436	3.0689	4.0393	5.4543
2012	1.5542	2.9506	4.3605	5.7529	8.2571	1.2657	2.2647	3.7293	4.6730	6.0382	1.1256	2.0429	3.2216	4.1534	5.5053
2013	1.6301	2.9930	4.3670	5.7963	8.1659	1.2793	2.3024	3.7585	4.7016	6.0352	1.1521	2.1150	3.3278	4.2748	5.6170
2014	1.5480	2.8989	4.2983	5.7087	8.3036	1.2804	2.2753	3.7590	4.6871	6.0865	1.1704	2.0846	3.3077	4.2710	5.7196
Mean	1.5425	2.9166	4.3054	5.7199	8.2967	1.2458	2.2439	3.7197	4.6551	6.0076	1.1242	2.0186	3.2181	4.1473	5.5508
Year	$CTE_{0.5}$	$CTE_{0.75}$	$CTE_{0.9}$	$CTE_{0.95}$	$CTE_{0.99}$	$CTE_{0.5}$	$CTE_{0.75}$	$CTE_{0.9}$	$CTE_{0.95}$	$CTE_{0.99}$	$CTE_{0.5}$	$CTE_{0.75}$	$CTE_{0.9}$	$CTE_{0.95}$	$CTE_{0.99}$
2009	3.4066	4.6149	6.2049	7.4209	9.5362	2.7420	3.7338	4.9597	5.5827	6.7183	2.3843	3.2635	4.4300	5.1468	6.4170
2010	3.3251	4.5053	6.0100	7.1432	9.1565	2.5864	3.5259	4.7032	5.3373	6.4478	2.2943	3.1102	4.1337	4.7695	5.9015
2011	3.2584	4.4294	5.9344	7.0777	9.0573	2.5614	3.4901	4.6883	5.3500	6.4696	2.2945	3.1229	4.1984	4.9024	6.0660
2012	3.3896	4.5664	6.0657	7.2016	9.1597	2.6877	3.6309	4.8437	5.4972	6.6082	2.3924	3.2473	4.3359	5.0201	6.1132
2013	3.4315	4.5871	6.0921	7.2088	9.2063	2.7222	3.6801	4.8689	5.5227	6.6011	2.4637	3.3412	4.4475	5.1345	6.2839
2014	3.3589	4.5151	6.0096	7.1461	9.2293	2.7125	3.6582	4.8731	5.5270	6.5365	2.4492	3.3204	4.4435	5.1599	6.3263
Mean	3.3617	4.5364	6.0528	7.1997	9.2242	2.6687	3.6198	4.8228	5.4695	6.5636	2.3797	3.2343	4.3315	5.0222	6.1847

[a] denotes $\times 10^3$

uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

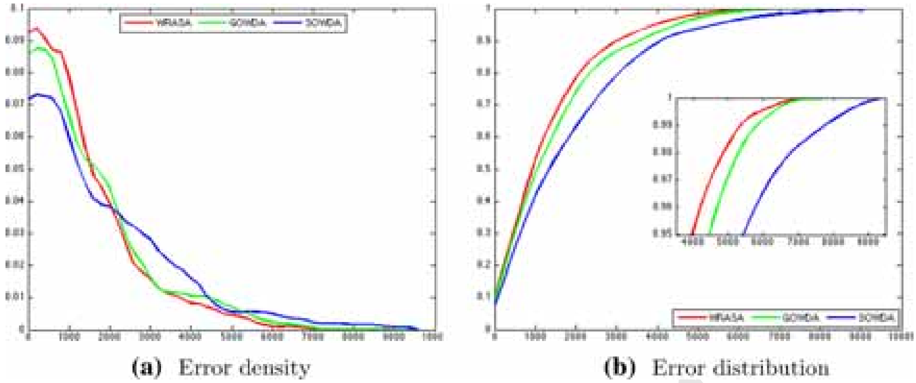


Fig. 4 Comparison of **a** error density and **b** error distribution in forecasting electricity power demand for three big data systems (i.e., SOWDA, GOWDA, and WRASA)

In order to graphically illustrate the performances we exhibit the complete plot of forecasting errors for the three big data systems in Fig. 4. Panel (a) in Fig. 4 presents the error density of the system' outcomes. We can see that all error densities are positively skewed to zero, but we cannot easily distinguish the difference. Panel (b) in Fig. 4 exhibits the error distribution where we can easily identify the quantile (i.e., TE) that is used to quantify the tail error. From the right panel of Fig. 4, we see the error distribution of WRASA is shifted to the left, which means that the performance of WRASA is better comparing with SOWDA and GOWDA. Table 4 illustrates this with concrete quantile values for $\alpha \geq 0.5$. With these values we can consistently compare the system performance after definite ranking TE and CTE values. We can thus verify that the proposed measures (i.e., TE and CTE) for evaluating quality perform well and can be confirmed to be very robust and efficient measures.

4.4 Discussion with simulation

In our empirical study of forecasting power consumption, all quality measurement tools we applied suggest to us the best performance in conformity. As we have seen in Fig. 3, the power consumption illustrates seasonality. Not surprisingly, WRASA works well for such a weak irregularity. In this case, the preternatural feature of the proposed measure (i.e., CTE) based on tail distribution is not startlingly clear. We are going to conduct a simulation study to investigate the performance of the proposed quality measure for assessing the performance of a BDS with strong irregularity.

4.4.1 Simulation method

In this study we perform Monte Carlo simulations where extreme irregular observations (jumps) are generated from two different patterns: (1) excessive volatility and (2) excessive volatility with regime switching. We generate 100 sequences of length 2^{12}. This trend is based on a sine function, whose amplitude and frequency are drawn from a uniform distribution. For generating the pattern (1) signals, following Sun and Meinl (2012), we add jumps to the trend. Jump occurrences are uniformly distributed (coinciding with a Poisson arrival rate observed in many systems), with the jump heights being a random number drawn from a skewed contaminated normal noise suggested by Chun et al. (2012). For generating pattern (2) data, we repeat the method used for pattern (2) signals, but shift the trend up and down

uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

once in order to characterize a sequence of excessive volatility with regime switching. The amplitude of the shift is four times the previous trend; see Sun et al. (2015).

Similar to how the method was applied in our empirical study, we also need to decide a moving window for the training and forecasting operations. We set the in-sample size as 200 and the out-of-sample size as 35, 80, and 100. The number of window moves is then 720, and we generate 100 data series for each pattern. Therefore, for each pattern we test our algorithm 72,000 times for in-sample approximation and one-step ahead forecasting. In our simulation, we have 216,000 runs in total. In our study, cases 1–3 and cases 4–6 are based on the data pattern (1) and (2), respectively. The out-of-sample forecasting length is set as 35 for cases 1 and 4, 80 for cases 2 and 5, and 100 for cases 3 and 6.

For each run we collect the values of six conventional criteria and five newly proposed measures. We report the results in next subsection.

4.4.2 Results and discussion

Table 5 reports the performance quality of three BDS evaluated by the six conventional methods. We can see that when comparing the performances, WRASA is considered superior by MAPE, MAE, and RMSE in Case 1 in training while SOWDA is preferred by CVM, KS, and Kuiper. Similarly, MAPE, MAE, and RMSE support WRASA and CVM, KS, and Kuiper support SOWDA for out-of-sample performance in Case 1. In Table 5 we highlight the best performance in bold. Obviously, we encounter difficulty in synchronizing because there are contradictive results.

The contradiction can be identified by four different classes. The first is noncoincidence among three mean error based methods, (i.e., MAPE, MAE, and RMSE). For example, for the out-of-sample performance of Case 3, GOWDA is preferred by RMSE whereas MAE and MAPE support WRASA. Second, there is noncoincidence among three extreme error based methods, (i.e., CVM, KS, and Kuiper). For example, for the in-sample performance of Case 4, SOWDA is preferred by CVM whereas KS and Kuiper support WRASA. Third, there is noncoincidence between the mean error based methods and extreme error based methods. For example, for the out-of-sample performance of Case 2, all mean error based methods support GOWDA but all extreme error based methods benefit WRASA. Fourth, one method is not distinguishable from another method. For example, for the out-of-sample performance of Case 5, MAE cannot distinguish which one is better between GOWDA and WRASA because a tie occurs in comparison. KS cannot distinguish SOWDA and WRASA when comparing the in-sample performance of Case 6.

When we apply TE and CTE, it turns out to be much easier. In Table 6, when we evaluate the in-sample performance, all smallest values of TE and CTE lead to WRASA. Therefore, its in-sample performance is better than other two. When we compare the out-of-sample performance, we recognize that TE cannot always provide a coherent conclusion with the results reported in Table 7. For example, for Case 2, $TE_{0.95}$ supports GOWDA whereas $TE_{0.90}$ benefit WRASA. Meantime, $TE_{0.99}$ cannot distinguish[5] the quality of GOWDA and WRASA. Not surprisingly, CTE can provide the coherent conclusion that WRASA always performs better. We have shown that CTE is a dynamic coherent measure since it satisfies all properties given by Axiom 3.

There are many advantages of TE. First, it is a simple quality measure and has a straightforward interpretation. Second, it defines the complete error distribution by specifying TE for all α. Third, it focuses only on the part of the error distribution specified by α. Fourth,

[5] Not because of making a round number.

uncorrected proof

Table 5 Comparison of the in-sample (modelling) and out-of-sample (forecasting) performances measured by six criteria for three big data systems (i.e., SOWDA, GOWDA, and WRASA) with the simulated data

	In-sample						Out-of-sample					
	MAPE[d]	MAE[b]	RMSE[a]	CVM[b]	KS[b]	Kuiper[b]	MAPE[c]	MAE	RMSE	CVM	KS[b]	Kuiper[a]
Case 1												
SOWDA	2.8609	6.0428	0.9947	**1.0379**	**1.2090**	**2.2025**	3.1413	0.6648	0.9646	**0.3903**	**0.5565**	**0.9533**
GOWDA	2.7896	5.8902	1.4135	1.2691	1.2500	2.2670	3.0504	0.6456	0.9461	0.4086	0.5718	0.9827
WRASA	**2.5445**	**5.3772**	**0.8999**	1.1001	1.2390	2.2385	**3.0351**	**0.6424**	**0.9438**	0.4059	0.5752	0.9889
Case 2												
SOWDA	2.8298	5.9772	0.9879	**1.0270**	**1.1965**	**2.1810**	4.6658	0.9860	1.3610	2.3552	1.1846	1.9354
GOWDA	2.7454	5.7989	1.3855	1.2090	1.2530	2.2695	4.6378	0.9800	1.3562	**2.3538**	**1.1634**	**1.8953**
WRASA	**2.5232**	**5.3307**	**0.8929**	1.1082	1.2570	2.2550	**4.6254**	**0.9773**	**1.3549**	2.3602	1.1704	1.9139
Case 3												
SOWDA	2.8590	6.0052	1.0208	**0.9138**	1.1650	2.1025	5.2998	1.1138	1.5161	2.8918	1.2516	2.0565
GOWDA	2.8863	6.0630	1.4504	1.0890	1.1950	2.1435	5.2420	1.1016	**1.5072**	**2.8406**	**1.2502**	**2.0351**
WRASA	**2.6128**	**5.4872**	**0.9213**	0.9882	**1.1585**	**2.1010**	**5.2401**	**1.1012**	1.5073	2.8597	1.2614	2.0576
Case 4												
SOWDA	5.2062	10.9598	1.5097	**8.3170**	2.4130	3.0445	3.7299	0.7862	1.0837	**0.4759**	**0.5983**	**1.0334**
GOWDA	4.5318	9.5418	1.6574	8.7453	2.4515	3.0080	3.6405	0.7673	1.0650	0.4889	0.6036	1.0379
WRASA	**4.4446**	**9.3581**	**1.3021**	8.6965	**2.4075**	**2.9670**	**3.6288**	**0.7648**	**1.0631**	0.4899	0.6142	1.0509
Case 5												
SOWDA	5.1722	10.8892	1.4991	**8.3586**	**2.3820**	3.0010	5.1520	1.0854	1.4590	**2.5185**	**1.1732**	**1.9521**
GOWDA	4.4960	9.4653	1.6314	8.7791	2.4465	**2.9910**	5.1181	**1.0782**	1.4527	2.6139	1.2070	1.9974
WRASA	**4.4507**	**9.3720**	**1.2990**	8.6072	2.4330	2.9990	**5.1176**	**1.0782**	**1.4519**	2.6037	1.1984	2.0014
Case 6												
SOWDA	5.1509	10.8440	1.4838	**8.3519**	**2.3965**	3.0235	5.3946	1.1364	1.5276	2.7675	**1.2846**	**2.0928**
GOWDA	4.4693	9.4102	1.6120	8.6625	2.4400	2.9860	5.3528	1.1277	1.5226	2.7958	1.2989	2.1087
WRASA	**4.4493**	**9.3683**	**1.2994**	8.5844	**2.3965**	**2.9565**	**5.3423**	**1.1254**	**1.5193**	**2.7475**	1.2876	2.0967

[a], [b], [c], and [d] denote $\times 10^{-1}$, $\times 10^{-2}$, $\times 10^{-3}$, and $\times 10^{-4}$ respectively

uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

Table 6 Quality evaluation with TE and CTE at different confidence levels for in-sample (training) performances of three big data systems (i.e., SOWDA, GOWDA, and WRASA) with the simulated data

	In-sample[a]					In-sample[a]				
	$TE_{0.50}$	$TE_{0.75}$	$TE_{0.90}$	$TE_{0.95}$	$TE_{0.99}$	$CTE_{0.50}$	$CTE_{0.75}$	$CTE_{0.90}$	$CTE_{0.95}$	$CTE_{0.99}$
Case 1										
SOWDA	0.6799	1.1943	1.8204	2.4232	6.7708	1.6068	2.2979	3.5520	5.0277	10.4363
GOWDA	0.5940	1.0297	1.5231	1.9042	3.6204	1.2454	1.6960	2.3725	3.0513	6.1440
WRASA	**0.5151**	**0.9094**	**1.3651**	**1.7223**	**3.5773**	**1.1323**	**1.5696**	**2.2645**	**3.0021**	**5.7232**
Case 2										
SOWDA	1.1386	1.5751	2.2680	3.0041	6.1230	1.5971	2.2790	3.5115	4.9499	10.2339
GOWDA	0.5913	1.0280	1.5226	1.8918	3.5748	1.2407	1.6889	2.3576	3.0261	6.1230
WRASA	**0.5220**	**0.9141**	**1.3724**	**1.7298**	**3.5321**	**1.1386**	**1.5751**	**2.2680**	**3.0041**	**5.6890**
Case 3										
SOWDA	0.6855	1.2028	1.8561	2.5256	7.1467	1.6442	2.3666	3.6996	5.2630	10.6361
GOWDA	0.5872	1.0208	1.5091	1.8872	3.8229	1.2414	1.6958	2.3891	3.1630	6.5599
WRASA	**0.5243**	**0.9203**	**1.3890**	**1.7574**	**3.7560**	**1.1618**	**1.6169**	**2.3590**	**3.0993**	**5.9082**
Case 4										
SOWDA	1.1815	2.0537	3.0416	3.8101	7.4555	2.4849	3.3867	4.7422	6.0995	11.0469
GOWDA	1.0731	1.8510	2.7011	3.2801	4.8327	2.1172	2.8008	3.6694	4.3751	6.9591
WRASA	**0.9073**	**1.5848**	**2.3596**	**2.9260**	**4.6202**	**1.8667**	**2.5141**	**3.3980**	**4.1808**	**6.6096**
Case 5										
SOWDA	1.1863	2.0604	3.0493	3.8182	7.3980	2.4854	3.3810	4.7198	6.0468	10.8072
GOWDA	1.0751	1.8516	2.7001	3.2738	4.8124	2.1173	2.8006	3.6646	4.3683	6.9476
WRASA	**0.9165**	**1.5987**	**2.3699**	**2.9216**	**4.6319**	**1.8745**	**2.5191**	**3.3956**	**4.1719**	**6.5948**
Case 6										
SOWDA	1.1835	2.0532	3.0421	3.7938	7.3004	2.4751	3.3653	4.6857	5.9938	10.6652
GOWDA	1.0761	1.8501	2.6987	3.2771	4.8358	2.1170	2.8008	3.6666	4.3736	6.7719
WRASA	**0.9151**	**1.5938**	**2.3663**	**2.9104**	**4.5879**	**1.8667**	**2.5057**	**3.3695**	**4.1260**	**6.5969**

[a] denotes $\times 10^{-1}$

uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

Table 7 Quality evaluation with TE and CTE at different confidence levels for out-of-sample (forecasting) performances of three big data systems (i.e., SOWDA, GOWDA, and WRASA) with the simulated data

	Out-of-sample					Out-of-sample				
	$TE_{0.50}$	$TE_{0.75}$	$TE_{0.90}$	$TE_{0.95}$	$TE_{0.99}$	$CTE_{0.50}$	$CTE_{0.75}$	$CTE_{0.90}$	$CTE_{0.95}$	$CTE_{0.99}$
Case 1										
SOWDA	0.4997	0.8934	1.2971	1.8829	3.6543	1.0998	1.5153	2.1916	2.8565	4.1205
GOWDA	0.4767	0.8732	1.2716	1.8254	3.6390	1.0749	1.4879	2.1549	2.8234	4.1036
WRASA	**0.4746**	**0.8670**	**1.2624**	**1.8180**	**3.6373**	**1.0698**	**1.4819**	**2.1499**	**2.8231**	**4.1015**
Case 2										
SOWDA	0.7504	1.3397	2.1330	2.9630	4.2837	1.6291	2.2333	3.1110	3.7346	4.5810
GOWDA	0.7436	1.3335	2.1276	**2.9551**	**4.2806**	2.2407	2.2407	3.1160	3.7402	4.5839
WRASA	**0.7388**	**1.3327**	**2.1240**	2.9601	**4.2806**	**1.6267**	**2.2326**	**3.1101**	**3.7339**	**4.5806**
Case 3										
SOWDA	0.8525	1.5219	2.4831	**3.2664**	4.5814	2.5224	2.5224	3.4307	4.0294	4.7852
GOWDA	0.8355	**1.5024**	**2.4679**	3.2666	4.5826	1.8320	2.5132	3.4297	4.0328	4.7870
WRASA	**0.8332**	1.5040	2.4719	3.2715	**4.5690**	**1.8317**	**2.5111**	**3.4292**	**4.0293**	**4.7741**
Case 4										
SOWDA	0.6205	1.0802	1.5386	2.0920	3.7333	1.7237	1.7237	2.3923	3.0251	4.2306
GOWDA	0.5981	1.0590	**1.5143**	**2.0509**	**3.7052**	1.2561	1.6967	2.3585	2.9884	4.2017
WRASA	**0.5954**	**1.0517**	1.5160	2.0542	3.7080	**1.2528**	**1.6950**	**2.3566**	**2.9873**	**4.1962**
Case 5										
SOWDA	0.8566	1.4935	2.2958	3.0761	4.4892	2.3937	2.3937	3.2746	3.8909	4.7536
GOWDA	0.8493	**1.4868**	**2.2752**	**3.0535**	4.4892	1.7665	2.3855	3.2626	3.8817	4.7532
WRASA	**0.8475**	1.4881	2.2790	3.0597	**4.4877**	**1.7656**	**2.3823**	**3.2592**	**3.8806**	**4.7492**
Case 6										
SOWDA	0.8884	1.5726	2.4578	3.2160	4.5947	2.5236	2.5236	3.4186	4.0388	4.8445
GOWDA	0.8780	1.5577	2.4405	3.2206	4.6037	1.8569	2.5161	3.4188	4.0445	4.8539
WRASA	**0.8775**	**1.5559**	**2.4337**	**3.2129**	**4.5809**	**1.8529**	**2.5090**	**3.4115**	**4.0292**	**4.8283**

uncorrected proof

Journal: **10479** Article No.: **2795** TYPESET DISK LE CP Disp.:**2018/2/14** Pages: **30** Layout: **Small**

it can be estimated with parametric models once the error distribution can be clarified and the estimation procedures are stable with different methods. However, quality control using TE may lead to undesirable results for skewed error distributions. It does not consider the properties of the error distribution beyond α. In addition, it is non-convex and discontinuous for discrete error distributions.

The advantages of applying CTE are threefold. First, it has superior mathematical properties, as we have shown, and preserves some properties of TE, such as easy interpretation and complete definition of error distribution with all α. Second, it is continuous with respect to α and coherent. Third, when making decision, we can optimize it with convex programming since $\mathrm{CTE}_{\alpha}(\omega_1\mathcal{K}_1 + \cdots + \omega_n\mathcal{K}_n)$ is a convex function with respect to weights $\omega_1, \ldots, \omega_n$ where $\sum_1^n \omega_n = 1$. However, challenges still remain when estimating CTE as it is heavily affected by tail modeling.

Coherence matters when we conduct system engineering either by reconfiguring systems or aggregating tasks. System engineering refers to simple configurations, such as parallelization, and complex ones, such as centralization, fragmentation, or distribution, as we show in Fig. 1. TE is not coherent because it does not satisfy Axiom 2-(2) in Sect. 2.3.1. Axiom 2-(2) states that if we combine two independent tasks, then the total error of them is not greater than the sum of the error associated with each of them. When we reconfigure systems, we try to avoid creating any extra errors (or system instability), but no one can guarantee it. In our empirical study, all three big data systems apply only plain configuration because the input data are simple high-frequency time series and work on a single task of forecasting the load. Therefore, we can apply TE and CTE for dynamical measuring following Definition 3.

Accuracy is given by systemic computability. Many computational methods can be used to accurately describe the probability distribution of errors; therefore, we can obtain a definite value for TE and CTE. Compared with conventional methods, with more suitable models or computational methods, we could optimally reduce the incapability of tail measures. If we refer to robustness as consistency and efficiency as comparability, then we can confidentially apply CTE, since it is a preferred method for quality control due to its robustness and efficiency. Therefore, the only thing we need for conducting the quality control is to choose or search for a proper α that reflects our own utility.

5 Conclusion

In this article we highlighted the challenge of big data systems and introduced some background information focusing on their quality management. We looked to describe the prevailing paradigm for quality management of these systems. We proposed a new axiomatic framework for quality management of a big data system (BDS) based on the tail error (TE) and conditional tail error (CTE) of the system. From a decision analysis perspective, particularly for the CTE, a dynamic coherent mechanism can be applied to perform these quality errors by continuously updating the new information through out the whole systemic processing. In addition, the time invariance property of the conditional tail error has been well defined to ensure the inherent characteristics of a BDS are able to satisfy quality requirements.

We applied the proposed methods with three big data systems (i.e., SOWDA, GOWDA, and WRASA) based on wavelet algorithms and conducted an empirical study to evaluate their performances with big electricity demand data from France through analysis and forecasting. Our empirical results confirmed the efficiency and robustness of the method we proposed herein. In order to fully illustrate the advantage of the proposed method, we also provided a simulation study to highlight some features that were not obtained from our

uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

empirical study. We believe our results will enrich the future applications in quality management.

As we have seen, the three big data systems we investigated in this article are mainly from fragmented (decentralized) systems where the interconnection of each component is still hierarchical in algorithms. When the underlying big data system exhibits strong distributed features—for example, there are tremendous interoperations—how to execute the dynamic coherent quality measure we proposed in this article should be paid more attention. In addition, some robust tests focusing on validating its conditional correlation over time should be proposed for future study.

References

Agarwal, R., Green, R., Brown, P., Tan, H., & Randhawa, K. (2013). Determinants of quality management practices: An empirical study of New Zealand manufacturing firms. *International Journal of Production Economics*, *142*, 130–145.

Artzner, P., Delbaen, F., Eber, J., Heath, D., & Ku, K. (2007). Coherent multiperiod risk adjusted values and Bellman's principle. *Annals of Operations Research*, *152*, 5–22.

Baucells, M., & Borgonovo, E. (2013). Invariant probabilistic sensitivity analysis. *Management Science*, *59*(11), 2536–2549.

Bion-Nadal, J. (2008). Dynamic risk measures: Time consistency and risk measures from BMO martingales. *Finance and Stochastics*, *12*(2), 219–244.

Bion-Nadal, J. (2009). Time consistent dynamic risk processes. *Stochastic Processes and their Applications*, *119*(2), 633–654.

Chen, Y., & Sun, E. (2015). *Jump detection and noise separation with singular wavelet method for high-frequency data*. Working paper of KEDGE BS.

Chen, Y., & Sun, E. (2018). Chapter 8: Automated business analytics for artificial intelligence in big data @x 4.0 era. In M. Dehmer & F. Emmert-Streib (Eds.), *Frontiers in Data Science* (pp. 223–251). Boca Raton: CRC Press.

Chen, Y., Sun, E., & Yu, M. (2015). Improving model performance with the integrated wavelet denoising method. *Studies in Nonlinear Dynamics and Econometrics*, *19*(4), 445–467.

Chen, Y., Sun, E., & Yu, M. (2017). Risk assessment with wavelet feature engineering for high-frequency portfolio trading. *Computational Economics*. https://doi.org/10.1007/s10614-017-9711-7.

Cheridito, P., & Stadje, M. (2009). Time-inconsistency of VaR and time-consistent alternatives. *Finance Research Letters*, *6*, 40–46.

Chun, S., Shapiro, A., & Uryasev, S. (2012). Conditional value-at-risk and average value-at-risk: Estimation and asymptotics. *Operations Research*, *60*(4), 739–756.

David, H., & Nagaraja, H. (2003). *Order statistics* (3rd ed.). Hoboken: Wiley.

Deichmann, J., Roggendorf, M., & Wee, D. (2015). McKinsey quarterly november: Preparing IT systems and organizations for the Internet of Things. 5

Hazen, B., Boone, C., Ezell, J., & Jones-Farmer, J. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, *154*, 72–80.

Keating, C., & Katina, P. (2011). Systems of systems engineering: Prospects and challenges for the emerging field. *International Journal of System of Systems Engineering*, *2*(2/3), 234–256.

Liu, Y., Muppala, J., Veeraraghavan, M., Lin, D., & Hamdi, M. (2013). *Data center networks: Topologies architechtures and fault-tolerance characteristics*. Berlin: Springer.

Maier, M. (1998). Architecting principles for systems-of-systems. *Systems Engineering*, *1*(4), 267–284.

Mellat-Parst, M., & Digman, L. (2008). Learning: The interface of quality management and strategic alliances. *International Journal of Production Economics*, *114*, 820–829.

O'Neill, P., Sohal, A., & Teng, W. (2015). Quality management approaches and their impact on firms' financial performance—An Australian study. *International Journal of Production Economics*. https://doi.org/10.1016/j.ijpe.2015.07.015i.

Parast, M., & Adams, S. (2012). Corporate social responsibility, benchmarking, and organizational performance in the petroleum industry: A quality management perspective. *International Journal of Production Economics*, *139*, 447–458.

Pham, H. (2006). *System software reliability*. Berlin: Springer.

Uncorrected proof

Journal: 10479 Article No.: 2795 TYPESET DISK LE CP Disp.:2018/2/14 Pages: 30 Layout: Small

Riedel, F. (2004). Dynamic coherent risk measures. *Stochastic Processes and their Applications*, *112*(2), 185–200.

Shooman, M. (2002). *Reliability of computer systems and networks: Fault tolerance analysis and design*. Hoboken: Wiley.

Sun, E., Chen, Y., & Yu, M. (2015). Generalized optimal wavelet decomposing algorithm for big financial data. *International Journal of Production Economics*, *165*, 161–177.

Sun, E., & Meinl, T. (2012). A new wavelet-based denoising algorithm for high-frequency financial data mining. *European Journal of Operational Research*, *217*, 589–599.

Sun, W., Rachev, S., & Fabozzi, F. (2007). Fractals or I.I.D.: Evidence of long-range dependence and heavy tailedness from modeling German equity market returns. *Journal of Economics and Business*, *59*, 575–595.

Sun, W., Rachev, S., & Fabozzi, F. (2009). A new approach for using Lèvy processes for determining high-frequency value-at-risk predictions. *European Financial Management*, *15*(2), 340–361.

Wu, S., & Zhang, D. (2013). Analyzing the effectiveness of quality management practices in China. *International Journal of Production Economics*, *144*, 281–289.

uncorrected proof

Journal: **10479** Article No.: **2795** TYPESET DISK LE CP Disp.:**2018/2/14** Pages: **30** Layout: **Small**

Journal: 10479
Article: 2795

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details required	Author's response
1.	Journal instruction requires a city and country for affiliations; however, these are missing in affiliations 1 and 2. Please verify if the provided city and country are correct and amend if necessary.	
2.	Please confirm the section headings are correctly identified.	
3.	Kindly check and confirm the layout of all the tables.	
4.	Please provide a definition for the significance of bold in the Tables 5, 6 and 7.	
5.	Kindly provide complete information for the reference Deichmann et al. (2015).	

uncorrected proof