Time consistent expected mean-variance in multistage stochastic quadratic optimization: a model and a matheuristic

Unai Aldasoro · María Merino · Gloria Pérez

Received: date / Accepted: date

Abstract In this paper, we present a multistage time consistent Expected Conditional Risk Measure for minimizing a linear combination of the expected mean and the expected variance, so-called Expected Mean-Variance. The model is formulated as a multistage stochastic mixed-integer quadratic programming problem combining risk-sensitive cost and scenario analysis approaches. The proposed problem is solved by a matheuristic based on the Branch-and-Fix Coordination method. The multistage scenario cluster primal decomposition framework is extended to deal with large-scale quadratic optimization by means of stage-wise reformulation techniques. A specific case study in risk-sensitive production planning is used to illustrate that a remarkable decrease in the expected variance (risk cost) is obtained. A competitive behavior on the part of our methodology in terms of solution quality and computation time is shown when comparing with plain use of CPLEX in 150 benchmark instances, ranging up to 711845 constraints and 193000 binary variables.

Keywords Branch-and-Fix Coordination · expected conditional risk measures · matheuristic algorithms · McCormick relaxation · Multistage stochastic optimization · quadratic mixed 0-1 programming · production planning · time consistent risk aversion

Mathematics Subject Classification (2000) 90C15 · 90C20 · 90C59

1 Introduction

Mathematical optimization is actually one of the most reliable tools for decision-making. The need to incorporate uncertainty into mathematical programming models gave rise to the field of Stochastic Optimization, which enables the risk inherent in decision making due to uncertainty in the main parameters of the problem to be managed as far as possible. Multistage stochastic optimization is

This research has been partially supported by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund through project MTM2015-65317-P (MINECO/FEDER/EU); via BCAM Severo Ochoa excellence accreditation grant SEV-2013-0323; by the Bizkaia Talent and EC COFUND program, request AYD-000-280; by the Basque Government through the BERC 2014-2017 program and Grupo de Investigación IT-928-16; and by the University of the Basque Country UPV/EHU through its UFI BETS 2011 program.

U. Aldasoro
Basque Center for Applied Mathematics, BCAM. Bilbao, Basque Country, Spain.
Tel.: +34-94-6567842
Fax: +34-94-6567843
E-mail: ualdasoro@bcamath.org

M. Merino · G. Pérez
Dep. of Applied Mathematics and Statistics and Operations Research. University of the Basque Country, UPV/EHU.

a good tool for representing uncertainty. Such problems have a more complex information structure scenario than (frequently approximated) two-stage models. In the general formulation of a multistage stochastic model, decisions on each stage have to be made stage-wise without anticipation of future events, i.e. the so-called non-anticipativity constraints (NAC), see Wets (1974,1975) [43,44], must be satisfied. The problem is formulated by what is known as the Deterministic Equivalent Model (DEM), a term coined by Wets (1974,1975) [43,44]. Conventionally, special attention has been given to optimizing the DEM by optimizing expected value of the objective function over the set of scenarios. In 1994 Kall and Wallace gave an introduction scenario analysis (see [28]) and in 1997 Birge and Louveaux explained the main concepts on stochastic optimization via scenario tree analysis (see [9]).

In multistage stochastic optimization problems where the parameters are random variables with known probability distributions, a multistage Risk Neutral (RN) strategy is conventionally considered. In multistage RN approaches the optimization is performed in average, i.e. the setting does not hedge against the occurrence of low probability high-consequence events. Alternatively, a decision-maker that is very concerned about non-desired scenarios may consider a risk-averse strategy for the optimization problem. Several risk averse measures were suggested by various authors, see some surveys in [35] and [38]. Additionally, it also seems natural to consider that the optimal decisions at a certain node of the tree should not depend on scenarios that cannot happen in the future, i.e. the so-named time consistency principle stated by Shapiro in 2009 (see [40]).

From a modeling perspective, in this paper we introduce a multistage time consistent risk averse measure, so-called Expected Mean-Variance (EMV), that can be included in the class of Expected Conditional Risk Measures (ECRM) proposed by Homem-de-Mello and Pagnoncelli in 2016 (see [32]). Interestingly, ECRMs have suitable properties for multistage scenario cluster primal decomposition frameworks. The proposed EMV model minimizes a linear combination of the expected mean and the expected variance, based on a combination of risk-sensitive cost (see [11]) and scenario analysis approaches. Consequently, EMV is formulated as a multistage stochastic mixed-integer quadratic programming problem.

In recent years, Quadratic Programming (QP) problems attracted considerable attention both from modeling and algorithmic standpoints. Many applications of science and technology are by construction, or after reformulations, described as QP or Mixed-Integer Quadratic (MIQ) problems (see Furini et al. [23] for a summary of applications). A classical linearization strategy that promotes very concise mixed 0-1 linear representations of mixed 0-1 quadratic programs was put forward by Glover in 1975 (see [25]). In 1976 McCormick presented a general method for obtaining a global solution in factorable nonlinear programming problems (see [31]). More recently, in 2004 Adams, Forrester and Glover gave a linearization strategy for mixed 0-1 quadratic problems that produces small formulations with tight relaxations by using binary identities to rewrite the objective (see [1]). In 2013 Kolodziej, Castro and Grossmann presented relaxation techniques for solving nonconvex bilinear programs (see [29]).

A number of recent papers have considered Stochastic Mixed Integer Quadratic (SMIQ) framework for modeling decision problems under uncertainty. To account for volatility in the power industry, in 2004 Conejo et al. focused on risk modeling, emphasizing the tradeoff between maximum profit and minimum risk as an MIQ (see [13]). In 2006 S. Ahmed proposed stochastic programming decomposition algorithms for mean-risk functions, see [2]. In 2008 Osorio, Gulpinar and Rustem presented a case study of a mixed integer stochastic programming approach to mean-variance post-tax portfolio management using scenario trees as an SMIQ model (see [36,37]). In 2009 Dentcheva and Ruszczyński presented multivariate stochastic dominance constraints and robust stochastic dominance for risk modeling, see [15,16]. Recently in 2017 Sun et al. introduced quadratic two-stage stochastic optimization with coherent measures of risk, see [42].

As it has been observed in the literature, decomposition algorithms are able to exploit the nice structure of models based on scenario analysis and convexity. Those approaches include, among others, the following types for two-stage and multistage problems: Benders Decomposition, Lagrangean Decomposition, Regularization, Progressive Hedging, Stochastic Dynamic Programming and scenario cluster primal decomposition (see [5] and references therein). Since 2005, Escudero et al. have developed a general algorithm based on a Branch-and-Fix Coordination (BFC) scenario cluster primal decomposition scheme for solving multistage stochastic mixed 0-1 problems (see [18, 19, 20], among others). Currently, it is possible to solve large-scale multistage stochastic mixed 0-1 linear problems by using BFC decomposition and parallelization techniques. In 2013 and 2017, Aldasoro et al. presented a parallel BFC algorithm for solving multistage stochastic mixed 0-1 problems, considering both exact (see [4]) and matheuristic approaches (see [5]).

Decomposition methods have also been explored for solving QP problems. In 1980 decomposition methods for solving multistage stochastic problems with recourse, with discrete distribution, quadratic objective function and linear inequality constraints were formulated by Louveaux (see [30]). In 2011, Birge and Louveaux presented a quadratic nested decomposition method for solving multistage stochastic convex quadratic models with discrete distribution (see [9]). In 2013 Cesarone et al presented a method for mixed integer quadratic programming in the context of mean-variance portfolio optimization, see [12]. In 2014 Siddiqui, Gabriel and Azarm presented an efficient optimization method that is applicable to mixed integer problems with quasiconvex constraints and convex objective function using Benders decomposition (see [41]). In 2015 Mijangos published an algorithm to solve two-stage stochastic nonlinear convex problems, based on the Branch-and-Fix-Coordination methodology of [17] for two-stage stochastic mixed 0-1 first-stage problems, see [33].

From an algorithmic perspective, in this paper we introduce a matheuristic decomposition algorithm for solving multistage stochastic convex quadratic 0-1 problems with bounded continuous variables, named the Quadratic matHeuristic Branch-and-Fix Coordination algorithm (QH-BFC). We generalize to quadratic optimization the BFC and branching approaches presented in 2017 by Aldasoro et al. (see [5]). The multistage scenario cluster primal decomposition framework is extended to deal with large-scale quadratic optimization by means of stage-wise reformulation techniques, rewriting the objective function and updating the set of constraints. The algorithm proposed can be applied to a wide variety of realistic problems, for example the stochastic portfolio selection problem, many process engineering problems, production planning problems, and scheduling problems, among others.

A risk-sensitive production planning is used as a specific case study (see in [39] the modeling, classification and reformulation of production planning problems and a review of models for production planning under uncertainty in [34]). We report a broad computational experience conducted to assess the solution quality of the matheuristic solution for 150 medium-scale and large-scale instances up to 711845 constraints and 193000 0-1 variables. The multistage stochastic mixed 0-1 quadratic instances are available in our SMIQLib repository [6]. The results evidence a remarkable decrease in absolute and relative risk costs, with an increase in expected cost.

The rest of the paper is organized as follows: Section 2 presents the description of the multistage stochastic mixed 0-1 quadratic problem and defines the time consistent EMV model. The reformulation of the DEM model using Fortet and Glover inequalities and the relaxation of the DEM model using the McCormick scheme are also described in this section. Section 3 details clustering partitioning and explains the matheuristic quadratic QH-BFC algorithm. Section 4 reports a description of a general multistage stochastic quadratic production planning problem and the main results of a broad computational experience to assess the validity of the QH-BFC algorithm for solving medium and large scale mixed 0-1 quadratic problems. The computation results are compared with the plain use of CPLEX v12.6.3. Section 5 concludes and outlines future work.

2 Time consistent EMV risk aversion in multistage stochastic quadratic optimization

2.1 Multistage stochastic mixed 0-1 quadratic problem

Consider the following multistage deterministic mixed 0-1 convex quadratic model:

$$
\begin{array}{lll}
\min \sum_{t\in\mathcal{T}} a_t x_t + b_t y_t + \frac{1}{2}\sum_{t,t'\in\mathcal{T}} (x_t, y_t, x_{t'}, y_{t'}) D_{tt'} \begin{pmatrix} x_t \\ y_t \\ x_{t'} \\ y_{t'} \end{pmatrix} & & \\
s.t. \ \sum_{\tau\in\mathcal{T}^t} (A^t_\tau x_\tau + B^t_\tau y_\tau) = h_t & \forall t\in\mathcal{T} & \\
x_t \in \{0,1\}^{nx_t}, \ 0 \le y_t \le M_t & \forall t\in\mathcal{T}, &
\end{array} \tag{1}
$$

where $\mathcal{T}$ is the set of stages, such that $T = |\mathcal{T}|$, x_t and y_t are the nx_t and ny_t dimensional vectors of the 0-1 and continuous variables, respectively, a_t and b_t are the row vectors of the objective function coefficients, respectively, M_t is the upper bound vector of the continuous variables, $D_{tt'}$ is a known positive semidefinite matrix of dimensions $(nx_t + ny_t + nx_{t'} + ny_{t'}) \times (nx_t + ny_t + nx_{t'} + ny_{t'})$, A^t_τ and B^t_τ are the $m \times nx_t$ and $m \times ny_t$ constraint matrices, respectively, for stages in $\mathcal{T}^t = \{1, \ldots, t\}$ and h_t is the right-hand-side vector (rhs) for stage $t \in \mathcal{T}$.

Using a scenario analysis approach, model (1) can be extended to consider uncertainty in any of the main parameters, namely, the objective function, rhs and constraint matrix coefficients. Let $(\Omega, \mathcal{F}, P)$ denote a probability space. Ω denotes the finite set of scenarios, where $\omega \in \Omega$ represents a specific scenario and w^ω denotes the probability P assigned by the modeler to scenario ω, such that $\sum_{\omega\in\Omega} w^\omega = 1$. Two scenarios are said to belong to the same group in a given stage when they have the same realizations of the uncertain parameters up to that stage. Following the *non-anticipativity* (NA) principle stated in [44], both scenarios should have the same value for the variables with time indexes up to the given stage. Let $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots \subset \mathcal{F}_T$ be sub sigma-algebras of $\mathcal{F}$, where $\mathcal{F}_t$ corresponds to the information available until stage t. A *multi-period* risk function F is a mappping from $\mathcal{Z}_1 \times \ldots \times \mathcal{Z}_T$ to $I\!R$, where $\mathcal{Z}_t$ denotes a space of $\mathcal{F}_t$-measurable functions from Ω to $I\!R$.

For the general formulation of a multistage model consider a multistage scenario tree to represent uncertainty, where $\mathcal{G}$ is the set of nodes. Let $g \in \mathcal{G}$ denote a node in the scenario tree that has a one-to-one correspondence with a scenario group, say g, in the same stage t, where $\mathcal{G}^t \subset \mathcal{G}$ denotes the set of nodes (and thus the related scenario groups) in stage $t = t(g)$, for $t \in \mathcal{T}$. Let $\tilde{\mathcal{A}}^g$ and $\mathcal{S}^g$ denote the sets of ancestor and successor nodes of node g (including itself), respectively. Also let $\Omega^g \subset \Omega$ denote the set of scenarios that belong to group g, such that $\Omega^1 = \Omega$, where $g = 1$ is the root node in the scenario tree, and Ω^g is singleton for $g \in \mathcal{G}^T$ (for that case, let $g = \omega$).

By slightly abusing the notation, the following deterministic quadratic problem emerges for $g \in \mathcal{G}^T$,

$$
\begin{array}{l}
z^g_{QP} = \min \sum_{q\in\mathcal{A}^g} a^q x^q + b^q y^q + \frac{1}{2}\sum_{q,q'\in\tilde{\mathcal{A}}^g} (x^q, y^q, x^{q'}, y^{q'}) D^{qq'} \begin{pmatrix} x^q \\ y^q \\ x^{q'} \\ y^{q'} \end{pmatrix} \\
s.t. \ \sum_{q\in\mathcal{A}^g} (A^q_g x^q + B^q_g y^q) = h^g \\
x^g \in \{0,1\}^{n^g_x}, \quad 0 \le y^g \le M^g.
\end{array} \tag{2}
$$

Consider the *compact representation* of the Deterministic Equivalent Model (DEM) of an SMIQ problem that minimizes the expected value of the quadratic function over the set of scenarios Ω in the scenario tree:

$$
\begin{aligned}
(DEM): z^{DEM} = \min \sum_{g \in \mathcal{G}^T} w^g [\sum_{q \in \tilde{\mathcal{A}}^g} a^q x^q + b^q y^q + \frac{1}{2} \sum_{q,q' \in \tilde{\mathcal{A}}^g} (x^q, y^q, x^{q'}, y^{q'}) D^{qq'} \begin{pmatrix} x^q \\ y^q \\ x^{q'} \\ y^{q'} \end{pmatrix}] \\
s.t. \sum_{q \in \mathcal{A}^g} (A_g^q x^q + B_g^q y^q) = h^g \qquad \forall g \in \mathcal{G} \\
x^g \in \{0,1\}^{n_x^g}, \quad 0 \le y^g \le M^g \qquad \forall g \in \mathcal{G},
\end{aligned} \tag{3}
$$

where w^g is the probability of scenario g, for $g \in \mathcal{G}^T$ in the scenario tree (note that for each node $g \in \mathcal{G}$, the weight $w^g = \sum_{\omega \in \Omega^g} w^\omega$ is recovered); $\mathcal{A}^g \subseteq \tilde{\mathcal{A}}^g$ is the set of the ancestor nodes of node g (including itself) whose decisions have direct influence (i.e., have non zero elements) on the constraints for node g, where $\mathcal{A}^1 = \{1\}$; x^g and y^g are the vectors of the 0-1 and continuous variables for node g, respectively; M^g is the upper bound vector of the continuous y^g for $g \in \mathcal{G}$; a^g and b^g are the vectors of the objective function coefficients for the 0-1 and continuous variables, respectively; A_g^q and B_g^q are the constraint matrices of ancestor node $q \in \mathcal{A}^g$ in node g for the vectors x^q and y^q, respectively; h^g is rhs for node g; n_x^g and n_y^g are the number of 0-1 and continuous variables, respectively, for $g \in \mathcal{G}$, $n_x = \sum_{g \in \mathcal{G}} n_x^g$ and $n_y = \sum_{g \in \mathcal{G}} n_y^g$.

The structure of the matrix of the quadratic form of the objective function in (3) considers quadratic terms inside a node or between pair of nodes under the same scenario (intra-scenarios) but not under different scenarios (inter-scenarios), as follows,

	v^1	v^2	$\dots$	v^{m_2}	v^{m_2+1}	$\dots$	v^{m_3}	$\dots$	$v^{m_{T-1}+1}$	$\dots$	v^{m_T}
v^1	$D^{1,1}$	$D^{1,2}$	$\dots$	D^{1,m_2}	D^{1,m_2+1}		D^{1,m_3}	$\dots$	$D^{1,m_{T-1}+1}$		D^{1,m_T}
v^2	$D^{2,1}$	$D^{2,2}$			D^{2,m_2+1}	$\dots$					
$\vdots$	$\vdots$		$\ddots$								
v^{m_2}	$D^{m_2,1}$			D^{m_2,m_2}		$\dots$	D^{m_2,m_3}				
v^{m_2+1}	$D^{m_2+1,1}$	$D^{m_2+1,2}$			D^{m_2+1,m_2+1}						
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\ddots$					
v^{m_3}	$D^{m_3,1}$			D^{m_3,m_2}			D^{m_3,m_3}				
$\vdots$	$\vdots$							$\ddots$			
$v^{m_{T-1}+1}$	$D^{m_{T-1}+1,1}$								$D^{m_{T-1}+1,m_{T-1}+1}$		
$\vdots$	$\vdots$									$\ddots$	
v^{m_T}	$D^{m_T,1}$										D^{m_T,m_T}

where $v^q = (x^q, y^q)$ is a $n_x^q + n_y^q$ vector of decisions, $m_t = \sum_{\tau=1}^t |\mathcal{G}_\tau|$ and the $(n_x^q + n_y^q + n_x^{q'} + n_y^{q'}) \times (n_x^q + n_y^q + n_x^{q'} + n_y^{q'})$ matrices $D^{qq'}$ are defined as

$$
D^{qq'} = \begin{pmatrix} D_{x^q x^q} & D_{x^q y^q} & D_{x^q x^{q'}} & D_{x^q y^{q'}} \\ D_{y^q x^q} & D_{y^q y^q} & D_{y^q x^{q'}} & D_{y^q y^{q'}} \\ D_{x^{q'} x^q} & D_{x^{q'} y^q} & D_{x^{q'} x^{q'}} & D_{x^{q'} y^{q'}} \\ D_{y^{q'} x^q} & D_{y^{q'} y^q} & D_{y^{q'} x^{q'}} & D_{y^{q'} y^{q'}} \end{pmatrix},
$$

for $q, q' \in \tilde{\mathcal{A}}^g, \ g \in \mathcal{G}^T$ and $D^{qq'} = 0$ if $\nexists g \in \mathcal{G}^T : q, q' \in \tilde{\mathcal{A}}^g$. See an example in Figure 1.

2.2 Time consistent EMV in multistage mixed 0-1 quadratic modeling

In stochastic quadratic optimization, quadratic terms can appear as a natural consequence of the modeling, i.e. models of the type (2) $\forall g \in \mathcal{G}$ and the DEM (3) in compact representation. On the

	(x^1,y^1)	(x^2,y^2)	(x^3,y^3)	(x^4,y^4)	(x^5,y^5)	(x^6,y^6)	(x^7,y^7)
(x^1,y^1)	D^{11}	D^{12}	D^{13}	D^{14}	D^{15}	D^{16}	D^{17}
(x^2,y^2)	D^{21}	D^{22}		D^{24}	D^{25}		
(x^3,y^3)	D^{31}		D^{33}			D^{36}	D^{37}
(x^4,y^4)	D^{41}	D^{42}		D^{44}			
(x^5,y^5)	D^{51}	D^{52}			D^{55}		
(x^6,y^6)	D^{61}		D^{63}			D^{66}	
(x^7,y^7)	D^{71}		D^{73}				D^{77}

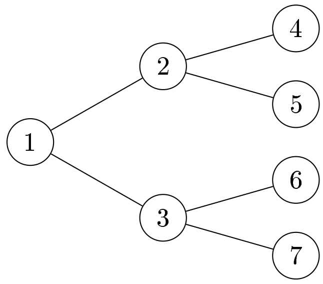

Fig. 1 Example of matrix structure (left) in 2^2 scenario tree (right)

other hand, quadratic terms can appear as the result of incorporating risk measures over a multistage Risk Neutral (RN) problem. In this subsection we detail the second way.

Given $\xi(\omega)$ a random vector with the discrete distribution $\xi(\omega) = (a_\omega, b_\omega, A_\omega, B_\omega, h_\omega, M_\omega)$ with likelihood w^ω for each scenario $\omega \in \Omega$, the corresponding values emerge in each group $g \in \mathcal{G} : \omega \in \Omega^g$, $\xi^g = (a^g, b^g, A^g, B^g, h^g, M^g)$ with likelihood w^g. ξ_t is a n_t-dimensional random vector representing the uncertainty observed in stage t, i.e., $\mathcal{F}_t$-measurable mapping from Ω to $I\!R^{n_t}$. We consider the following linear mixed 0-1 DEM in compact representation,

$$\begin{aligned} z^{MIP} = \min & \sum_{g\in\mathcal{G}} w^g[a^g x^g + b^g y^g] \\ s.t. & \sum_{q\in\mathcal{A}^g} (A^q_g x^q + B^q_g y^q) = h^g \qquad \forall g \in \mathcal{G} \\ & x^g \in \{0,1\}^{n^g_x}, \quad 0 \le y^g \le M^g \quad \forall g \in \mathcal{G}, \end{aligned} \tag{4}$$

where a^g, b^g coefficients are random vectors in the objective function with known distribution functions for all $g \in \mathcal{G}$ (see Figure 2).

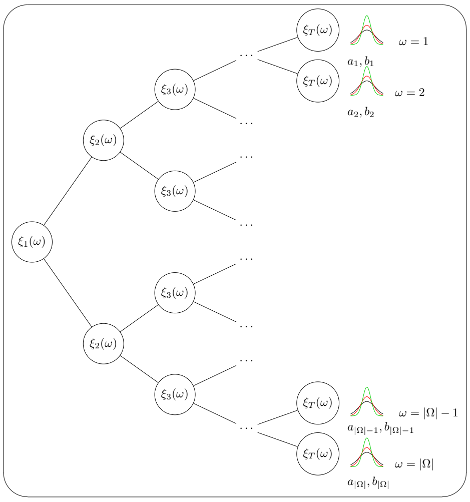

Fig. 2 Example of random vectors a and b over a multistage scenario tree.

For that reason, consider the following random vectors, where expectations, variances, and covariances can be calculated based on their distribution functions.

a^g: is the random cost for the $x-$variable in node $g \in \mathcal{G}$, with expectation $\mu_{a^g} := E(a^g)$ and covariance matrix $\Sigma_{a^g a^g} := Var(a^g) = Cov(a^g, a^g)$ of dimension $n_x^g \times n_x^g$.

b^g: is the random cost for the $y-$variable in node $g \in \mathcal{G}$, with expectation $\mu_{b^g} := E(b^g)$ and covariance matrix $\Sigma_{b^g b^g} := Var(b^g) = Cov(b^g, b^g)$ of dimension $n_y^g \times n_y^g$ and mixed covariance $\Sigma_{a^g b^g} := Cov(a^g, b^g)$ of dimension $n_x^g \times n_y^g$.

In RN stochastic optimization, the objective function considers the expectation over the set of scenarios. Typically, this corresponds to the maximization of the expected returns or the minimization of the expected loss in financial and industrial applications, respectively.

$$\begin{aligned} z^{RN} &= \min \sum_{g \in \mathcal{G}} w^g [\mu_{a^g} x^g + \mu_{b^g} y^g] \\ s.t. & \sum_{q \in \mathcal{A}^g} (A_g^q x^q + B_g^q y^q) = h^g \qquad \forall g \in \mathcal{G} \\ & x^g \in \{0,1\}^{n_x^g}, \quad 0 \le y^g \le M^g \qquad \forall g \in \mathcal{G}. \end{aligned} \tag{5}$$

Instead of minimizing only the expected loss (the conventional goal in an RN environment), we propose also to minimize the variability in terms of variances and covariances.

The cost for group $g \in \mathcal{G}$ is given by $z^g = \sum_{q \in \tilde{\mathcal{A}}^g} (a^q x^q + b^q y^q)$ where $\mathcal{A}^g$ is the set of indexes for the ancestor nodes of group g. The expectation cost for group $g \in \mathcal{G}^T$ under (a,b)-distribution is

$$E(z^g) = \sum_{q \in \tilde{\mathcal{A}}^g} (\mu_{a^q} x^q + \mu_{b^q} y^q) \tag{6}$$

and the variance cost for group $g \in \mathcal{G}^T$ under (a,b)-distribution is

$$Var(z^g) = \sum_{q,q' \in \tilde{\mathcal{A}}^g} (x^q y^q x^{q'} y^{q'}) D^{qq'} \begin{pmatrix} x^q \\ y^q \\ x^{q'} \\ y^{q'} \end{pmatrix}, \tag{7}$$

where the $(n_x^q + n_y^q + n_x^{q'} + n_y^{q'}) \times (n_x^q + n_y^q + n_x^{q'} + n_y^{q'})$ matrices $D^{qq'}$ are defined as follows

$$D^{qq'} = \begin{pmatrix} \Sigma_{a^q a^q} & \Sigma_{a^q b^q} & \Sigma_{a^q a^{q'}} & \Sigma_{a^q b^{q'}} \\ \Sigma_{b^q a^q} & \Sigma_{b^q b^q} & \Sigma_{b^q a^{q'}} & \Sigma_{b^q b^{q'}} \\ \Sigma_{a^{q'} a^q} & \Sigma_{a^{q'} b^q} & \Sigma_{a^{q'} a^{q'}} & \Sigma_{a^{q'} b^{q'}} \\ \Sigma_{b^{q'} a^q} & \Sigma_{b^{q'} b^q} & \Sigma_{b^{q'} a^{q'}} & \Sigma_{b^{q'} b^{q'}} \end{pmatrix}.$$

Remember that the covariance matrix is symmetric semidefinite positive. If the distribution functions are not fully known, the expectations, variances and covariance matrices can be estimated using good estimators. At this point, the following possible objective functions can be considered:

1. The expected conditional mean of the cost for all groups g, $g \in \mathcal{G}^T$, based on (6):

$$\mu = \sum_{g \in \mathcal{G}^T} w^g E(z^g). \tag{8}$$

   In this case, the multistage stochastic DEM model (5) is an RN model.
2. The expected conditional variance of the cost for all groups g, $g \in \mathcal{G}^T$, based on (7):

$$\sigma^2 = \sum_{g \in \mathcal{G}^T} w^g Var(z^g). \tag{9}$$

   In this case, the multistage stochastic DEM model is a variance model.

3. The β linear combination of both (8) and (9), where $\beta > 0$. For each group $g \in \mathcal{G}^T$ the risk-sensitive objective function is given by $E(z^g) + \beta \cdot Var(z^g)$ where β is the *risk-aversion* parameter (see definition of risk-sensitive cost in [11]). By extension of that definition, the objective function function can be considered over all groups g, $g \in \mathcal{G}^T$, based on (6) and (7):

$$\mu + \beta \cdot \sigma^2 = \sum_{g \in \mathcal{G}^T} w^g [E(z^g) + \beta \cdot Var(z^g)]. \tag{10}$$

We called to this risk function, *Expected Mean-Variance (EMV)*. We propose a multistage risk-sensitive stochastic mixed 0-1 problem defining the objective function in the DEM in compact representation as in (10), in an extension of the risk-sensitive function to multistage stochastic models. Since a decrease in expected variance produces an increase in expected cost, so the decision-maker should carefully establish the value of the risk-aversion parameter β. As an example, a risk conservative decision-maker places more emphasis on minimizing risk and therefore defines a large value of β to increase the weight of the risk measure.

Using the objective function EMV, the Deterministic Equivalent Model is as follows,

$$\begin{array}{lll} z_{EMV}^{DEM} = \min \sum\limits_{g \in \mathcal{G}^T} w^g \sum\limits_{q \in \tilde{\mathcal{A}}^g} [E(z^q) + \beta \cdot Var(z^q)] & & \\ \quad s.t. \sum\limits_{q \in \mathcal{A}^g} (A_g^q x^q + B_g^q y^q) = h^g & \forall g \in \mathcal{G} & \\ \quad x^g \in \{0,1\}^{n_x^g}, \quad 0 \le y^g \le M^g & \forall g \in \mathcal{G}. & \end{array} \tag{11}$$

Note that we have defined two multistage stochastic quadratic problems (3) and (11) in compact representation with the same structure. Model (3) is a multistage risk-sensitive model (13) when the coefficients of the linear terms a^q and b^q are the expectations μ_{a^q} and μ_{b^q} and each matrix $D^{qq'}$ of the quadratic terms is the matrix $2\beta Cov(a^q, b^q, a^{q'}, b^{q'})$ for groups $q, q' \in \tilde{\mathcal{A}}^g$, $g \in \mathcal{G}^T$. For this reason, without loss generality, from now we consider the multistage stochastic quadratic problem in compact representation (3).

Our notion of *consistency* is inspired on definitions available in literature, see [32,40]. Time consistency concept states that optimal decisions should not depend on scenario tree branches that we know at stage t that cannot happen in later stages. Consider the problem $z_{EMV_{g'}}^{DEM}$ at a given node $g' \in \mathcal{G}^t$ as follows: optimizing z_{EMV}^{DEM} (11) when all the information and decisions from previous stages are known, i.e:

$$\begin{array}{lll} z_{EMV_{g'}}^{DEM} = \min \sum\limits_{g \in \mathcal{G}^T \cap \mathcal{S}^{g'}} w^g \sum\limits_{q \in \tilde{\mathcal{A}}^g} [E(z^q) + \beta \cdot Var(z^q)] & & \\ \quad s.t. \sum\limits_{q \in \mathcal{A}^g} (A_g^q x^q + B_g^q y^q) = h^g & \forall g \in \mathcal{G} \cap (\mathcal{S}^{g'} \cup \tilde{\mathcal{A}}^{g'}) & \\ \quad x^g = \dot{x}^g, \quad y^g = \dot{y}^g & \forall g \in \mathcal{G} \cap \tilde{\mathcal{A}}^{g'} \; : \; t(g) < t & \\ \quad x^g \in \{0,1\}^{n_x^g}, \quad 0 \le y^g \le M^g & \forall g \in \mathcal{G} \cap \mathcal{S}^{g'} \; : \; t(g) \ge t. & \end{array} \tag{12}$$

where $(\dot{x}^g, \; \dot{y}^g)$ is an optimal solution in (11).

Definition 1 We say that the EMV risk measure is *time consistent* for problem z_{EMV}^{DEM} if there exists an optimal solution (x^*, y^*) of (11) such that for any stage $t \in \mathcal{T}$ and any node $g \in \mathcal{G}^t$, coincides with an optimal solution of problem (12), and this is true for any data and scenario tree.

That is, the optimality of our decisions today remain optimal if we solve the problem tomorrow with the known information updated.

Proposition 1 *The Expected Mean Variance model (11) is time consistent.*

Proof For any data, any scenario tree, any stage $t \in \mathcal{T}$ and any node $g \in \mathcal{G}^t$, if we fix (x, y) decisions to the optimal solution of (11), say $(\dot{x}, \dot{y})$, until stage $t-1$ to the model (12), the solution $(\dot{x}^g, \dot{y}^g)$ where $g \in \mathcal{G} \ : \ t(g) \geq t$ is feasible in (12). Let us prove the optimality by *reductio ad absurdum*. Consider $((\hat{x}^g, \hat{y}^g))_{t(g)\geq t}$ an optimal solution of (12) which is better than $((\dot{x}^g, \dot{y}^g))_{t(g)\geq t}$. As decisions of previous stages are fixed and therefore, non-anticipativity is satisfied, it is also feasible solution of (11) and its objective function would be better than the optimal one (contradiction). The result follows because the model (11) where the decisions until stage t are fixed, is separable in $|\mathcal{G}^t|$ independent submodels as (12), where each submodel is attached to a subtree rooted in $g \in \mathcal{G}^t$. □

Let us consider the class of risk measures defined in [32]:

Definition 2 The family of the *Expected Conditional Risk Measures (ECRM)* is a class of multi-period risk measures F defined as follows,

$$F(Z_1, \ldots, Z_T) = Z_1 + \rho_2(Z_2) + E_{\xi_{[2]}}[\rho_3^{\xi_{[2]}}(Z_3)] + E_{\xi_{[3]}}[\rho_4^{\xi_{[3]}}(Z_4)] + \ldots + E_{\xi_{[T-1]}}[\rho_T^{\xi_{[T-1]}}(Z_T)],$$

where $\xi_{[t]}$ is a realization $\xi_1, \ldots, \xi_t$, E indicates the expectation with respect to the corresponding variables and ρ_t is a one-period risk measure applied to period t.

Proposition 2 *The EMV (11) is a Expected Conditional Risk Measure.*

Proof It results because the model (11) can be described in terms of previous definition. Note that the objective function is as follows $[E(z^1) + \beta \cdot Var(z^1)] + \sum_{g\in\mathcal{G}^2} w^g[E(z^g) + \beta \cdot Var(z^g)] + \ldots + \sum_{g\in\mathcal{G}^T} w^g[E(z^g) + \beta \cdot Var(z^g)]$, where $w^g = \sum_{\omega\in\Omega^g} w^\omega \quad \forall g \in \mathcal{G}$. □

2.3 Reformulation of quadratic mixed 0-1 models

It is frequently possible to, explicitly tighten the bounds of the continuous variables y^g, $g \in \mathcal{G}$. From now on, let $\underline{y}^g$ and $\overline{y}^g$ be the lower and upper bound vectors of y^g, respectively, in each group $g \in \mathcal{G}$, so $\underline{y}^g \leq y^g \leq \overline{y}^g$. If it is not possible to tighten the bounds for a variable y_j^g, $\underline{y}_j^g = 0$ and $\overline{y}_j^g = M_j^g$, $j \in \mathcal{I}_y^g$.

$$\begin{aligned}
(DEM): z^{DEM} = \min \sum_{g\in\mathcal{G}^T} w^g [\sum_{q\in\tilde{A}^g} a^q x^q + b^q y^q + \frac{1}{2}\sum_{q,q'\in\tilde{A}^g} (x^q, y^q, x^{q'}, y^{q'}) D^{qq'} \begin{pmatrix} x^q \\ y^q \\ x^{q'} \\ y^{q'} \end{pmatrix}] \\
s.t. \sum_{q\in\mathcal{A}^g} A_g^q x^q + B_g^q y^q = h^g \qquad \forall g \in \mathcal{G} \\
x^g \in \{0,1\}^{n_x^g}, \quad \underline{y}^g \leq y^g \leq \overline{y}^g \qquad \forall g \in \mathcal{G}.
\end{aligned} \tag{13}$$

We consider reformulation of the quadratic terms where the 0-1 variables are present with and without continuous ones. Let $\mathcal{I}_x^g$ and $\mathcal{I}_y^g$ denote the set of indexes of the variables in vectors x^g and y^g, respectively, for $g \in \mathcal{G}$.

Remark 1 For notation simplification, we will consider reformulation on product of variables under the same node $g \in \mathcal{G}$. By analogy, it can be consider for variables under different nodes $g, g' \in \mathcal{G}$.

In our DEM (13) the square terms on the 0-1 variables can be replaced by linear terms, $(x_i^g)^2 = x_i^g$, so that the coefficients of the linear terms x_i^g in the objective function are updated, such that

$a_i^g := \frac{1}{2}D_{ii}^g + (a^g)_i,\ i \in \mathcal{I}_x^g, g \in \mathcal{G}$. By using the Fortet inequalities scheme [22] the 0-1 quadratic terms $x_i^g x_j^g$ can be replaced by the 0-1 variables $\tilde{x}_{ij}^g$ in the objective function, whose linear coefficients are $c_{ij}^g = \frac{1}{2}D_{ij}^g,\ i,j \in \mathcal{I}_x^g,\ i<j,\ g \in \mathcal{G}$, plus the following Reformulation Constraint (RC) system:

$$\begin{array}{l} x_i^g + x_j^g \le 1 + \tilde{x}_{ij}^g, \quad i,j \in \mathcal{I}_x^g,\ i<j,\ g \in \mathcal{G} \\ \tilde{x}_{ij}^g \le x_i^g,\ \tilde{x}_{ij}^g \le x_j^g,\ i,j \in \mathcal{I}_x^g,\ i<j,\ g \in \mathcal{G}. \end{array} \tag{14}$$

By applying the scheme in Glover [25] to the mixed 0-1 quadratic terms, $x_i^g y_j^g$ can be replaced in the objective function by the continuous variables $\tilde{y}_{ij}^g$, whose coefficients are $d_{ij}^g = \frac{1}{2}D_{ij}^g$, $i \in \mathcal{I}_x^g,\ j \in \mathcal{I}_y^g,\ g \in \mathcal{G}$ and the following RC system is added:

$$\begin{array}{ll} \underline{y}_j^g x_i^g \le \tilde{y}_{ij}^g \le \overline{y}_j^g x_i^g & \forall i \in \mathcal{I}_x^g, j \in \mathcal{I}_y^g,\ g \in \mathcal{G} \\ y_j^g + (x_i^g - 1)\overline{y}_j^g \le \tilde{y}_{ij}^g \le y_j^g + (x_i^g - 1)\underline{y}_j^g & \forall i \in \mathcal{I}_x^g, j \in \mathcal{I}_y^g,\ g \in \mathcal{G} \\ \tilde{y}_{ij}^g \le y_j^g - (x_i^g - 1)\underline{y}_j^g & \forall i \in \mathcal{I}_x^g, j \in \mathcal{I}_y^g,\ g \in \mathcal{G}. \end{array} \tag{15}$$

Consequently, $D_{ij}^g = 0$, for $i,j \in \mathcal{I}_x^g$, $i \in \mathcal{I}_x^g$ and $j \in \mathcal{I}_y^g$ or $i \in \mathcal{I}_y^g$, and $j \in \mathcal{I}_x^g$, $g \in \mathcal{G}$.

We define the reformulated MIQ DEM (16) with only bilinear and square terms with continuous variables as follows,

$$\begin{array}{ll} z^{DEM} = \min \sum_{g \in \mathcal{G}^T} w^g [\sum_{q \in \tilde{A}^g} a^q x^q + b^q y^q + c^q \tilde{x}^q + d^q \tilde{y}^q + \frac{1}{2} \sum_{q,q' \in \tilde{A}^g} (y^q, y^{q'}) D_{yy}^{qq'} \begin{pmatrix} y^q \\ y^{q'} \end{pmatrix}] & \\ s.t. \sum_{q \in \mathcal{A}^g} A_g^q x^q + B_g^q y^q = h^g & \forall g \in \mathcal{G} \\ RC\ (14-15) & \\ x^g \in \{0,1\}^{n_x^g},\ \tilde{x}^g \in \{0,1\}^{n_{\tilde{x}}^g},\ \underline{y}^g \le y^g \le \overline{y}^g,\ \tilde{y}^g \le \overline{y}^g & \forall g \in \mathcal{G}, \end{array} \tag{16}$$

where the matrix D_{yy}^g is guaranteed to be semidefinite positive (see [24]).

2.4 Relaxed formulation of quadratic mixed 0-1 models

By using the McCormick scheme [29], it is possible to replace the continuous quadratic terms $y_i^g y_j^g$ where $\underline{y}_i^g \le y_i^g \le \overline{y}_i^g$ and $\underline{y}_j^g \le y_j^g \le \overline{y}_j^g$, by the continuous variable u_{ij}^g in model (16). The strict bilinear terms, $y_i^g y_j^g$ can be replaced in the objective function by the continuous variable u_{ij}^g whose coefficients are the elements $f_{ij}^g = D_{ij}^g$, $i \ne j$, $i,j \in \mathcal{I}_y^g$ of the matrix D_{yy}^g, $g \in \mathcal{G}$ and the quadratic terms $(y_i^g)^2$ can be replaced by u_{ii}^g, whose coefficients are the elements $f_{ii}^g = \frac{1}{2}D_{ii}^g$, $i \in \mathcal{I}_y^g$, of the matrix D_{yy}^g, $g \in \mathcal{G}$.

Remark 2 For notation simplification, we will describe relaxed formulation on product of continuous variables under the same node $g \in \mathcal{G}$. By analogy, it can be consider for variables under different nodes $g, g' \in \mathcal{G}$.

Definition 3 A ***McCormick stage*** $m^* \in \mathcal{T}$ is the stage until the McCormick scheme [29] is applied in multistage stochastic continuous quadratic models.

If m^* is the *McCormick* stage, a relaxed model of (16) can be generated. If the bilinear terms of the continuous variables $y_i^g y_j^g$ are replaced in model DEM (16) by u_{ij}^g, $i,j \in \mathcal{I}_y^g$, $g \in \mathcal{G}^t$ of the stages

t, $t \le m^*$ and the corresponding McCormick Constraints (McCC) (17) are added then the relaxed model (19) is obtained.

$$\begin{aligned}
&u^g_{ij} \ge y^g_i \underline{y}^g_j + \underline{y}^g_i y^g_j - \underline{y}^g_i \underline{y}^g_j && \forall i,j \in \mathcal{I}^g_y,\ g \in \mathcal{G}^t,\ t \le m^* \\
&u^g_{ij} \ge y^g_i \overline{y}^g_j + \overline{y}^g_i y^g_j - \overline{y}^g_i \overline{y}^g_j && \forall i,j \in \mathcal{I}^g_y,\ g \in \mathcal{G}^t,\ t \le m^* \\
&u^g_{ij} \le y^g_i \underline{y}^g_j + \overline{y}^g_i y^g_j - \overline{y}^g_i \underline{y}^g_j && \forall i,j \in \mathcal{I}^g_y,\ g \in \mathcal{G}^t,\ t \le m^* \\
&u^g_{ij} \le y^g_i \overline{y}^g_j + \underline{y}^g_i y^g_j - \underline{y}^g_i \overline{y}^g_j && \forall i,j \in \mathcal{I}^g_y,\ g \in \mathcal{G}^t,\ t \le m^*.
\end{aligned} \tag{17}$$

Note that for the quadratic terms, the McCC until stage m^* are as follows,

$$\begin{aligned}
&u^g_{ii} \ge 2\underline{y}^g_i y^g_i - (\underline{y}^g_i)^2 && \forall i \in \mathcal{I}^g_y,\ g \in \mathcal{G}^t,\ t \le m^* \\
&u^g_{ii} \ge 2\overline{y}^g_i y^g_i - (\overline{y}^g_i)^2 && \forall i \in \mathcal{I}^g_y,\ g \in \mathcal{G}^t,\ t \le m^* \\
&u^g_{ii} \le (\underline{y}^g_i + \overline{y}^g_i) y^g_i - \underline{y}^g_i \overline{y}^g_i && \forall i \in \mathcal{I}^g_y,\ g \in \mathcal{G}^t,\ t \le m^*.
\end{aligned} \tag{18}$$

By slightly abusing the notation, the relaxed model can be expressed as follows,

$$\begin{aligned}
&\underline{z} = \min \sum_{g \in \mathcal{G}^T} w^g [\sum_{q \in \tilde{A}^g} a^q x^q + b^q y^q + c^q \tilde{x}^q + d^q \tilde{y}^q + \sum_{q \in \bigcup_{t \le m^*} \mathcal{G}^t \cap \tilde{A}^g} f^q u^q + \frac{1}{2} \sum_{q,q' \in \bigcup_{t > m^*} \mathcal{G}^t \cap \tilde{A}^g} (y^q, y^{q'}) D^{qq'}_{yy} \begin{pmatrix} y^q \\ y^{q'} \end{pmatrix}] \\
&\quad s.t.\ Constraints\ system\ in\ model\ (16) \\
&\qquad McC\ (17).
\end{aligned} \tag{19}$$

Note that if $m^* = T$ the relaxed model (19) is a linear mixed 0-1 problem.

It is possible to find lower bounding formulations for solving bilinear programming problem (16), considering the McCormick convex envelopes [29,31] for bounding the bilinear terms and solving model (19). The optimal solution $\underline{z}$ of the relaxed problem (19) provides a lower bound of the optimal solution of the DEM (16). Furthermore, if the coefficient $D^g_{ij} = f^g_{ij} > 0$ then $u^g_{ij} \le y^g_i y^g_j$, therefore $f^g_{ij} u^g_{ij} \le D^g_{ij} y^g_i y^g_j$, $g \in \mathcal{G}^t$, $t \le m^*$ and if $D^g_{ij} = f^g_{ij} < 0$ then $u^g_{ij} \ge y^g_i y^g_j$, therefore $f^g_{ij} u^g_{ij} \le D^g_{ij} y^g_i y^g_j$, $g \in \mathcal{G}^t$, $t \le m^*$.

Furthermore, an upper bound $\overline{z}$ of the optimal solution of the DEM (16) can be obtained by computing the objective function of the problem (16) fixing the values of the common variables obtained from the optimal solution of the relaxed problem (19). Feasibility is obviously guaranteed and $\underline{z} \le z^{DEM} \le \overline{z}$.

Note that if the variable y^g_i, $i \in \mathcal{I}^g_y$ is semi-continuous (see [45]), the constraint system of model (16) includes the following constraint $\underline{y}^g_i x^g_i \le y^g_i \le \overline{y}^g_i x^g_i$. If $x^g_i = 0$ then $y_i = 0$ and each bilinearity $y^g_i y^g_j = 0$, $\forall j \in \mathcal{I}^g_y$.

3 Matheuristic algorithm for multistage stochastic quadratic optimization

3.1 Cluster partitioning

We now decompose the scenario tree into a set of scenario cluster subtrees, each for a different scenario cluster in the set denoted as $\mathcal{C} = \{1, ..., C\}$ with $C = |\mathcal{C}|$. Let Ω_c denote the set of scenarios that belongs to cluster c, such that $\Omega_c \bigcap \Omega_{c'} = \emptyset$, $c, c' \in \mathcal{C} : c \ne c'$ and $\Omega = \cup_{c \in \mathcal{C}} \Omega_c$.

We propose to choose the number of scenario clusters C as any value from the subset $\{|\mathcal{G}_1|, |\mathcal{G}_2|, \ldots, |\mathcal{G}_T|\}$. Accordingly, consider the following definitions taken from [19,20].

Definition 4 A ***break stage*** $t^* \in \mathcal{T}$ is a stage such that the number of scenario clusters is $C = |\mathcal{G}_{t^*+1}|$, where $t^*+1 \in \mathcal{T}$. In this case, any cluster $c \in \mathcal{C}$ is induced by a group, say g_c, for $g_c \in \mathcal{G}_{t^*+1}$ and contains all scenarios belonging to that group, i.e., $\Omega_c = \Omega^{g_c}$.

Notice that the choice of $t^* = 0$ corresponds to the full DEM and $t^* = T-1$ corresponds to the full scenario partitioning.

Assume that the scenario set is broken down into C scenario clusters. The cluster submodels and the full DEM (13) are to be formulated via a mixture of the *splitting variable* and *compact representations*, so that the submodels are linked by the explicit NAC up to break stage t^* (see below). Additionally, let $\mathcal{G}_c \subset \mathcal{G}$ denote the node set of the subtree that supports scenario submodel c, such that $\Omega^g \cap \Omega_c \neq \emptyset$ means that $g \in \mathcal{G}_c$, and $\mathcal{G}_c^t = \mathcal{G}^t \cap \mathcal{G}_c$ is the node set in stage $t \in \mathcal{T}$ in the subtree supporting cluster submodel $c \in \mathcal{C}$. The MIQ submodel for c can be expressed (in compact representation) as follows:

$$
\begin{aligned}
z_c = \min & \sum_{g \in \mathcal{G}_c \cap \mathcal{G}^T} w_c^g \Big[\sum_{q \in \tilde{A}^g} a^q x_c^q + b^q y_c^q + \frac{1}{2} \sum_{q,q' \in \tilde{A}^g} (x_c^q, y_c^q, x_c^{q'}, y_c^{q'}) D^{qq'} \begin{pmatrix} x_c^q \\ y_c^q \\ x_c^{q'} \\ y_c^{q'} \end{pmatrix} \Big] \\
s.t. & \sum_{q \in \mathcal{A}^g} A_g^q x_c^q + B_g^q y_c^q = h^g \qquad \forall g \in \mathcal{G}_c \\
& x_c^g \in \{0,1\}^{n_x^g}, \;\; \underline{y}_c^g \leq y_c^g \leq \overline{y}_c^g \qquad \forall g \in \mathcal{G}_c,
\end{aligned}
\tag{20}
$$

where $w_c^g = \sum_{\omega \in \Omega^g \cap \Omega_c} w^\omega$ for $g \in \mathcal{G}_c$ Now split the set of stages $\mathcal{T}$ into two subsets, such that $\mathcal{T} = \mathcal{T}_1 \bigcup \mathcal{T}_2$, where $\mathcal{T}_1 = \{1, \ldots, t^*\}$, and $\mathcal{T}_2 = \{t^*+1, \ldots, T\}$.

The full DEM (13) is formulated by a mixture of the *splitting variable representation* (to explicitly satisfy the NAC between the cluster submodels) and the *compact representation* (to implicitly satisfy the NAC of each cluster). So, the cluster *splitting-compact representation* can be expressed as follows:

$$
\begin{aligned}
z^{DEM} = \min & \sum_{c \in \mathcal{C}} \sum_{g \in \mathcal{G}_c \cap \mathcal{G}^T} w_c^g \Big[\sum_{q \in \tilde{A}^g} a^q x_c^q + b^q y_c^q + \frac{1}{2} \sum_{q,q' \in \tilde{A}^g} (x_c^q, y_c^q, x_c^{q'}, y_c^{q'}) D^{qq'} \begin{pmatrix} x_c^q \\ y_c^q \\ x_c^{q'} \\ y_c^{q'} \end{pmatrix} \Big] \\
s.t. & \sum_{q \in \mathcal{A}^g} A_g^q x_c^q + B_g^q y_c^q = h^g \qquad \forall g \in \mathcal{G}_c, \; c \in \mathcal{C} \\
& NAC \; (22-23) \\
& x_c^g \in \{0,1\}^{n_x^g}, \; \tilde{x}_c^g \in \{0,1\}^{n_{\tilde{x}}^g}, \; \underline{y}_c^g \leq y_c^g \leq \overline{y}_c^g, \; \tilde{y}_c^g \leq \overline{y}_c^g \qquad \forall g \in \mathcal{G}_c, \; c \in \mathcal{C}.
\end{aligned}
\tag{21}
$$

Now, in the splitting-compact representation of the DEM (21) the NA principle is implicitly taken into account for the stages $t \in \mathcal{T}_2$. On the other hand, the (explicit) NAC of the variables in the nodes that belong to the stages in set $\mathcal{T}_1$ can be formulated by observing that the clusters c and c' have the node g in common if $g \in \mathcal{G}_c \cap \mathcal{G}_{c'}$, and this can only happen for $g \in \mathcal{G}^t : t \in \mathcal{T}_1$. Let C^g denote the set of (lexicographically ordered) indexes of the scenario clusters to which node g belongs, for $g \in \mathcal{G}^t$, $t \in \mathcal{T}_1$, such that $n(c)$ is the element immediately next to element c, for $c \in C^g/\{|C^g|\}$, where $|C^g|$ is the last element in set C^g.

So, the cluster submodels (20) are linked by the NAC to be formulated as follows:

$$
x_c^g - x_{n(c)}^g = 0 \;\forall c \in C^g/\{|C^g|\}, g \in \mathcal{G}^t, t \in \mathcal{T}_1. \tag{22}
$$

$$
y_c^g - y_{n(c)}^g = 0 \;\forall c \in C^g/\{|C^g|\}, g \in \mathcal{G}^t, t \in \mathcal{T}_1. \tag{23}
$$

Given $\underline{y}_c^g$ and $\overline{y}_c^g$ the tightened lower and upper bounds of vectors y_c^g, for each cluster $c \in \mathcal{C}^g$, it is possible to join the tightened bounds associated with replicated variables y_c^g, $c \in C^g, g \in \mathcal{G}^t, t \in \mathcal{T}_1$, in the splitting-compact representation of model (21). The bounds can be updated as follows:

$$\underline{y}_c^g := \max_{c \in C^g} \{\underline{y}_c^g\} \qquad \overline{y}_c^g = \min_{c \in C^g} \{\overline{y}_c^g\} \quad \forall g \in \mathcal{G}^t, t \in \mathcal{T}_1. \tag{24}$$

The reformulated cluster *splitting-compact representation* of model (16) can then be expressed as follows:

$$\begin{aligned} z^{DEM} = \min \sum_{c \in \mathcal{C}} \sum_{g \in \mathcal{G}_c \cap \mathcal{G}^T} w_c^g [\sum_{q \in \tilde{A}^g} a^q x_c^q + b^q y_c^q + c^q \tilde{x}_c^q + d^q \tilde{y}_c^q + \frac{1}{2} \sum_{q,q' \in \tilde{A}^g} (y_c^q, y_c^{q'}) D_{yy}^{qq'} \begin{pmatrix} y_c^q \\ y_c^{q'} \end{pmatrix}] \\ s.t. \ \textit{Constraints system in model } (21) \\ RC \ (14-15). \end{aligned} \tag{25}$$

Note that it is not necessary to add the NAC for the $\tilde{x}$-variables and the $\tilde{y}$-variables because if the x-variables and the y-variables satisfy the NAC then the $\tilde{x}$-variables and the $\tilde{y}$-variables satisfy the NAC.

The reformulated scenario cluster MIQ submodel c, for $c \in \mathcal{C}$, can be expressed as follows:

$$\begin{aligned} z_c = \min \sum_{g \in \mathcal{G}_c \cap \mathcal{G}^T} w_c^g [\sum_{q \in \tilde{A}^g} a^q x_c^q + b^q y_c^q + c^q \tilde{x}_c^q + d^q \tilde{y}_c^q + \frac{1}{2} \sum_{q,q' \in \tilde{A}^g} (y_c^q, y_c^{q'}) D_{yy}^{qq'} \begin{pmatrix} y_c^q \\ y_c^{q'} \end{pmatrix}] \\ s.t. \ \textit{Constraints system in model } (20) \\ RC \ (14-15). \end{aligned} \tag{26}$$

If m^* is the McCormick stage, the corresponding McCC (27) in the splitting representation are as follows:

$$\begin{aligned} u_{ijc}^g &\geq y_{ic}^g \underline{y}_{jc}^g + \underline{y}_{ic}^g y_{jc}^g - \underline{y}_{ic}^g \underline{y}_{jc}^g && \forall i,j \in \mathcal{I}_y^g,\ c \in \mathcal{C}^g,\ g \in \mathcal{G}^t,\ t \leq m^* \\ u_{ijc}^g &\geq y_{ic}^g \overline{y}_{jc}^g + \overline{y}_{ic}^g y_{jc}^g - \overline{y}_{ic}^g \overline{y}_{jc}^g && \forall i,j \in \mathcal{I}_y^g,\ c \in \mathcal{C}^g,\ g \in \mathcal{G}^t,\ t \leq m^* \\ u_{ijc}^g &\leq y_{ic}^g \underline{y}_{jc}^g + \overline{y}_{ic}^g y_{jc}^g - \overline{y}_{ic}^g \underline{y}_{jc}^g && \forall i,j \in \mathcal{I}_y^g,\ c \in \mathcal{C}^g,\ g \in \mathcal{G}^t,\ t \leq m^* \\ u_{ijc}^g &\leq y_{ic}^g \overline{y}_{jc}^g + \underline{y}_{ic}^g y_{jc}^g - \underline{y}_{ic}^g \overline{y}_{jc}^g && \forall i,j \in \mathcal{I}_y^g,\ c \in \mathcal{C}^g,\ g \in \mathcal{G}^t,\ t \leq m^*. \end{aligned} \tag{27}$$

Then the relaxed cluster *splitting-compact representation* of model (25), replacing the bilinear terms of the continuous variables of the stages t with $t \leq m^*$ and adding the McCC (27) can be expressed as follows:

$$\begin{aligned} \underline{z} = \min \sum_{c \in \mathcal{C}} \sum_{g \in \mathcal{G}_c \cap \mathcal{G}^T} w_c^g [\sum_{q \in \tilde{A}^g} a^q x_c^q + b^q y_c^q + c^q \tilde{x}_c^q + d^q \tilde{y}_c^q + \sum_{q \in \bigcup_{t \leq m^*} \mathcal{G}^t \cap \tilde{A}^g} f^q u_c^q + \\ \frac{1}{2} \sum_{q,q' \in \bigcup_{t > m^*} \mathcal{G}^t \cap \tilde{A}^g} (y_c^q, y_c^{q'}) D_{yy}^{qq'} \begin{pmatrix} y_c^q \\ y_c^{q'} \end{pmatrix}] \\ s.t. \ \textit{Constraints system in model } (25) \\ McCC \ (27). \end{aligned} \tag{28}$$

Note that if $m^* = T$ the relaxed model (28) is a linear mixed 0-1 problem.

It is worth pointing out that the efficiency of an MIQ engine for solving models (25) and (28) is very low, but it paves the way for performing the appropriate model decomposition. In our QH-BFC algorithm (see Algorithm 1 in Section 3.3) we branch fixing x- variables to 0 or 1 (line 8). Note that if the model (28) has been solved we obtain a feasible solution of model (25). Therefore, we again obtain a lower bound, $\underline{z}$, and an upper bound, $\overline{z}$, of the objective function value, $\underline{z} \leq z^{DEM} \leq \overline{z}$.

3.2 Matheuristic BFC algorithm (QH-BFC)

The Branch-and-Fix Coordination (BFC) methodology is a multistage scenario cluster primal decomposition framework, which is very suitable for solving *Expected Conditional Risk Measures*. It is assumed in this section that the main concepts and definitions of the BFC are known (see [20]), as is the matheuristic version of a dynamically incomplete branching H-BFC algorithm (see [5]), so the concepts are used directly to present the models required for solving SMIQ problems.

An integer Twin Node Family is a Twin Node Family (TNF) such that all common variables have already been branched on or fixed to 0-1 values and their related $x-$NAC (22) are satisfied. The integer TNF models in the algorithm are of two types:

a) By extension of the BFC procedure to quadratic models (see [5]) we need to solve the model (QP), where the common variables have been fixed to their 0-1 variables,

$$(QP): \ z_{QP}^{TNF} = \min \sum_{c\in\mathcal{C}} \sum_{g\in\mathcal{G}_c\cap\mathcal{G}^T} w_c^g [\sum_{q\in\tilde{A}^g} a^q x_c^q + b^q y_c^q + c^q \tilde{x}_c^q + d^q \tilde{y}_c^q + \frac{1}{2}(y_c^q, y_c^{q'}) D_{yy}^{qq'} \begin{pmatrix} y_c^q \\ y_c^{q'} \end{pmatrix}] \quad (29)$$
$$\text{s.t. Constraint system in model (25)}$$
$$x_c^g = \dot{x}_c^g \quad \forall g \in \mathcal{G}_c, \ c \in \mathcal{C},$$

where $\dot{x}_c^g$ denotes the 0-1 value vector of vector x_c^g in scenario cluster submodel c, for $c \in \mathcal{C} : g \in \mathcal{G}_c$ and, $x_c^g = \dot{x}_c^g$ provided that the $x-$ NAC (22) are satisfied for $g \in \mathcal{G}^t$, $t \in \mathcal{T}_1$. Notice that the x-variables are fixed to 0 or 1.
If model (29) is feasible then the new incumbent solution value is $\overline{z}^{DEM} := \min\{z_{QP}^{TNF}, \ \overline{z}^{DEM}\}$.
Let $(\dot{x}^{g,TNF}, \ y^{g,TNF}, \tilde{x}^{g,TNF}, \tilde{y}^{g,TNF})$ denote the solution of model (29) $\forall g \in \mathcal{G}$. Observe that all the NAC are satisfied.
But, given a positive McCormick stage m^*, we propose to solve a relaxation of problem (29), i.e. the following continuous quadratic problem ($\underline{\text{QP}}$):

$$(\underline{QP}): \underline{z}_{QP}^{TNF} = \min \sum_{g\in\mathcal{G}\cap\mathcal{G}^T} w^g [\sum_{q\in\tilde{A}^g} a^q x^q + b^q y^q + c^q \tilde{x}^q + d^q \tilde{y}^q + \sum_{q\in \bigcup_{t\le m^*} \mathcal{G}^t\cap\tilde{A}^g} f^q u^q +$$
$$\frac{1}{2} \sum_{q,q'\in \bigcup_{t>m^*} \mathcal{G}^t\cap\tilde{A}^g} (y^q, y^{q'}) D_{yy}^{qq'} \begin{pmatrix} y^q \\ y^{q'} \end{pmatrix}] \quad (30)$$
$$\text{s.t.} \quad \text{Constraint system in model (28)}$$
$$x_c^g = \dot{x}_c^g \quad \forall g \in \mathcal{G}_c, \ c \in \mathcal{C},$$

and if $m^* = T$, problem (30) is a linear problem.
If model (30) is feasible we obtain a lower bound $\underline{z}_{QP}^{TNF}$ of model (29) and by computing the objective function of (29) with the optimal solution of (30) $(\underline{\dot{x}}^{g,TNF}, \ \underline{y}^{g,TNF}, \underline{\tilde{x}}^{g,TNF}, \underline{\tilde{y}}^{g,TNF})$, we can also obtain an upper bound, $\overline{z}_{QP}^{TNF}$ of model (29). Then, $\underline{z}_{QP}^{TNF} \le z_{QP}^{TNF} \le \overline{z}_{QP}^{TNF}$, and the new incumbent solution value is $\overline{z}^{DEM} := \min\{\overline{z}_{QP}^{TNF}, \ \overline{z}^{DEM}\}$.

b) By extension of the BFC procedure to quadratic models (see [5]), we need to solve the model (MIQ), where the vectors x^g, y^g, $\tilde{x}^g$ and $\tilde{y}^g$ are fixed to the values $\dot{x}^{g,TNF}$, $y^{g,TNF}$, $\tilde{x}^{g,TNF}$ and $\tilde{y}^{g,TNF}$, $\forall g \in \mathcal{G}^t$, $t \in \mathcal{T}_1$, obtained by solving model (29) if $m^* = 0$ and model (30) if $m^* > 0$.

$$(MIQ): \ z_f^{TNF} = \min \sum_{c\in\mathcal{C}} \sum_{g\in\mathcal{G}_c\cap\mathcal{G}^T} w_c^g [\sum_{q\in\tilde{A}^g} a^q x_c^q + b^q y_c^q + c^q \tilde{x}_c^q + d^q \tilde{y}_c^q + \frac{1}{2}(y_c^q, y_c^{q'}) D_{yy}^{qq'} \begin{pmatrix} y_c^q \\ y_c^{q'} \end{pmatrix}]$$
$$\text{s.t. Constraint system in model (25)} \quad (31)$$
$$x_c^g = \dot{x}_c^{g,TNF}, \ y_c^g = y_c^{g,TNF}, \ \tilde{x}_c^g = \tilde{x}_c^{g,TNF}, \ \tilde{y}_c^g = \tilde{y}_c^{g,TNF}$$
$$\forall g \in \mathcal{G}^t \cap \mathcal{G}_c, \ c \in \mathcal{C}, \ t \in \mathcal{T}_1.$$

Notice that the x-variables, y-variables, $\tilde{x}$-variables and $\tilde{y}$-variables are fixed for the stages in $\mathcal{T}_1$. Observe that model (31) is easily decomposed by scenario clusters, such that

$$z_f^{TNF} = \sum_{c\in\mathcal{C}} z_{f,c}^{TNF}, \tag{32}$$

where $z_{f,c}^{TNF}$ is the solution value of models (33) for $c \in \mathcal{C}$, where variables for the stages in $\mathcal{T}_1$ have been fixed according to the solutions in previous models (29) or (30).

$$\begin{array}{l} z_{f,c}^{TNF} = \min \sum_{g\in\mathcal{G}_c\cap\mathcal{G}^T} w_c^g [\sum_{q\in\tilde{A}^g} a^q x_c^q + b^q y_c^q + c^q \tilde{x}_c^q + d^q \tilde{y}_c^q + \frac{1}{2}(y_c^q, y_c^{q'}) D_{yy}^{qq'} \begin{pmatrix} y_c^q \\ y_c^{q'} \end{pmatrix}] \\ \quad s.t.\ Constraint\ system\ in\ model\ (26) \\ \quad\quad x_c^g = \hat{x}_c^{g,TNF},\ y_c^g = y_c^{g,TNF},\ \tilde{x}_c^g = \tilde{x}_c^{g,TNF},\ \tilde{y}_c^g = \tilde{y}_c^{g,TNF} \\ \quad\quad \forall g \in \mathcal{G}^t \cap \mathcal{G}_c,\ t \in \mathcal{T}_1. \end{array} \tag{33}$$

In summary, the main differences between the Quadratic matHeuristic BFC algorithm (QH-BFC) applied for mixed 0-1 quadratic problems and previous H-BFC procedure are the following:

1. The scenario cluster submodels (26) and (33) to be solved for any candidate TNF are mixed 0-1 quadratic problems
2. We introduce reformulation techniques for transforming an SMIQ problem into a particular SMIQ problem with square and strict bilinear terms only for the continuous variables.
3. We define the McCormick stage m^* to generate relaxed problems associated with the quadratic mixed 0-1 problems. Then in one step of the algorithm instead of solving TNF models (29) (in general, continuous quadratic problems) we can introduce the solving of the relaxed problems up to stage m^* (see (30)), replacing bilinear terms and adding the McCC(27) in its generation.
4. The solution of problem (30) provides lower and upper bounds of the objective function of the associated problem (29).

3.3 Procedure for solving stochastic quadratic problems

A QH-BFC algorithm is presented below for solving stochastic quadratic problems. Before solving stochastic mixed 0-1 quadratic problems using QH-BFC algorithm we order models and decompose them into clusters. These techniques are detailed in [4] and [20]. A rough description of the dynamically-guided branching scheme for any TNF consists of branching first on the i variable in the 0-1 $\delta_{\bar{i}}$ direction, which can be expressed as follows:

$$\delta_{\bar{i}} = \begin{cases} 0, \text{ if } & \sum_{c\in\mathcal{C}_{\bar{i}}} (\hat{x}_c)_{\bar{i}} \le \frac{1}{2}|\mathcal{C}_{\bar{i}}| \\ 1, \text{ otherwise}, & \end{cases}$$

where $(\hat{x}_c)_{\bar{i}}$ denotes the value of variable $(x_c)_{\bar{i}}$ in the solution of submodels (26) and $\mathcal{C}_{\bar{i}} \subset \mathcal{C}$ denotes the set of cluster submodels (26) that have variable $x_{\bar{i}}$ in common.

The algorithm is based on coordinated branching on 0-1 variables until break stage. Iteratively lower bounds (scenario cluster models) and feasible solutions (described in the previous section) are solved until the memory or time limit is exceeded or the stopping criterion is satisfied. Notice that the stopping criterion (SC) is $\left|\frac{\overline{z}_{\nu-1}^{DEM} - \overline{z}_{\nu}^{DEM}}{\overline{z}_{\nu}^{DEM}}\right| < \epsilon$, where ν is the actual incumbent solution number in the algorithm and ϵ is a small positive quantity.

Algorithm 1 QH-BFC quadratic matheuristic algorithm

Initialization: $\nu := 0$, $\overline{z}_{\nu}^{DEM} := \infty$, $\overline{i} := 0$, t^*, m^* and ϵ;
Generate the scenario cluster MIQ submodels (26) and (33);
Generate the auxiliary McCormick relaxed QP (30);
while (memory/time limit not exceeded **and** SC not satisfied) **do**
  Solve MIQ (26) and compute $z = \sum_{c \in \mathcal{C}} z_c$ and $(\delta_i)_{i \in \mathcal{I}_x}$;
  if ($z \geq \overline{z}_{\nu}^{DEM}$ **or** NAC (22) not satisfied) **then**
    Update TNF node $\overline{i}$ according to the dynamically guided branching scheme;
    Fix TNF node $\overline{i}$ according to $\delta_{\overline{i}}$ in the coordinated BF trees;
  else
    if NAC (23) not satisfied **then**
      Solve QP (30), compute $\overline{z}_{QP}^{TNF}$ and update $z_{QP}^{TNF} := \overline{z}_{QP}^{TNF}$;
      if $z_{QP}^{TNF} < \overline{z}_{\nu}^{DEM}$ **then**
        Update $\overline{z}_{\nu}^{DEM} = z_{QP}^{TNF}$, test SC and $\nu := \nu + 1$;
        Solve MIQ (33) and compute $z_f^{TNF} = \sum_{c \in \mathcal{C}} z_{fc}^{TNF}$;
        Update $\overline{z}_{\nu}^{DEM} = \min(\overline{z}_{\nu}^{DEM}, z_f^{TNF})$, test SC and $\nu := \nu + 1$;
      end if
    else
      Update $\overline{z}_{\nu}^{DEM} := z$, test SC and $\nu := \nu + 1$;
    end if
  end if
end while
Output: QH-BFC algorithm obtains its best known solution with value $\overline{z}_{\nu}^{DEM}$.

4 Computational experience

The computational experiments were conducted in the ARINA computational cluster from SGI/IZO-SGIker at UPV/EHU [8], which provides 2248 cores divided as follows: 2160 xeon cores, 88 Itanium2 cores, with RAM memory between 16 and 512 Gb per node. For these computational experiments servers with two Broadwell-EP processors (Intel Xeon CPU E5-2680 v4 @ 2.40GHz) were used, each processor having 14 cores, with hand hyperthreading being disabled. The nodes had 128Gb of RAM and a solid state hard drive. For each optimization problem 8 cores were used and the memory was limited to 100 Gb.

The QH-BFC algorithm was implemented in a `C++` experimental code which uses the CPLEX v12.6.3 optimizer with 8 threads, (see [27]). The optimizer is used by QH-BFC to solve the MIQ submodels (26) and (33) for the set of scenario clusters $\mathcal{C}$ in different steps, the QP submodels (30) and the MIQ submodels (29) and (31). For solving MIQ problems using the CPLEX optimizer, in 2014 Bliek, Bonami and Lodi discussed classical algorithmic approaches, their implementation within CPLEX and new algorithmic advances (see [10]).

4.1 Pilot case: multistage stochastic EMV production planning problem

For computational experience, we consider an application in production planning and its extension to a multistage EMV model in stochastic quadratic optimization. A linear deterministic formulation for the general product structure capacitated multi-level lot-sizing model with lost demand is given in [39], a linear two stage stochastic formulation is given in [7], and a multistage stochastic formulation is given in [14]. Now, a multistage risk-sensitive cost production planning formulation is given for scheduling the production levels of a certain set of items in a planning horizon of several time periods.

Sets:

- $\mathcal{T}$, set of stages, $|\mathcal{T}| = T$.
- $\mathcal{G}$, set of scenario groups.
- $\mathcal{J}$, set of items to be produced, $|\mathcal{J}| = J$.
- $\mathcal{K}$, set of available shared resources, $|\mathcal{K}| = K$.

Deterministic parameters for item j, $j \in \mathcal{J}$:

- α_j^k, the amount of resource k consumption by one unit of item j, for $k \in \mathcal{K}$, $j \in \mathcal{J}$.
- γ_j^k, the amount of resource k for a set-up of item j, for $k \in \mathcal{K}$, $j \in \mathcal{J}$.

Stochastic parameters for item j in node g, $j \in \mathcal{J}, g \in \mathcal{G}$:

- $\underline{s}_{j,t(g)}$, safety stock of item j at the end of each period of node g.
- $\overline{s}_{j,t(g)}$, maximum volume in stock of item j at the end of each period of node g.
- d_j^g, the total demand of item j in node g.
- p_j^g, the fixed production cost of item j in node g.
- q_j^g, unit production cost of item j in node g.
- h_j^g, unit inventory holding cost of item j in group g.
- f_j^g, unit lost demand penalty of item j in node g.
- $\underline{y}_j^g$,$\overline{y}_j^g$, the bounds of production capacity of item j in node g.
- L_k^g, the available capacity of resource k in node g.

The variables for item j in node g, $j \in \mathcal{J}, g \in \mathcal{G}$:

- X_j^g, 0-1 variable whose value is 1 if item j is produced in node g.
- Y_j^g, semicontinuous variable, production of item j in node g.
- Z_j^g, continuous variable, stock of item j at the end of related time period in node g.
- F_j^g, continuous variable, lost demand of item j in node g.

We define the time consistent EMV production planning DEM with quadratic terms for the continuous Y-variables, Z-variables and F-variables in compact representation as follows,

$$
\begin{aligned}
z^{DEM} = \min \sum_{g \in \mathcal{G}^T} w^g [& \sum_{q \in \tilde{A}^g} \sum_{j \in \mathcal{J}} [\hat{\mu}_{p_j^q} X_j^q + \hat{\mu}_{q_j^q} Y_j^q + \hat{\mu}_{h_j^q} Z_j^q + \hat{\mu}_{f_j^q} F_j^q] + \\
& \beta \sum_{q,q' \in \tilde{A}^g} \sum_{i,j \in \mathcal{J}} (Y_i^q, Z_i^q, F_i^q, Y_i^{q'}, Z_i^{q'}, F_i^{q'}) \hat{D}_{ij}^{qq'} (Y_j^q, Z_j^q, F_j^q, Y_j^{q'}, Z_j^{q'}, F_j^{q'})^t]
\end{aligned}
$$

$s.t.$

$$
\begin{aligned}
& Z_j^g = Z_j^{\pi(g)} + Y_j^g - d_j^g + F_j^g && \forall j \in \mathcal{J},\ g \in \mathcal{G}. \\
& \underline{y}_j^g X_j^g \leq Y_j^g \leq \overline{y}_j^g X_j^g && \forall j \in \mathcal{J},\ g \in \mathcal{G} \\
& \underline{s}_{j,t(g)} \leq Z_j^g \leq \overline{s}_{j,t(g)} && \forall j \in \mathcal{J},\ g \in \mathcal{G} \\
& \sum_{j \in \mathcal{J}} \alpha_j^k Y_j^g + \sum_{j \in \mathcal{J}} \gamma_j^k X_j^g \leq L_k^g && \forall k \in \mathcal{K},\ g \in \mathcal{G} \\
& Y_j^g, Z_j^g, F_j^g \geq 0,\ X_j^g \in \{0,1\}, && \forall j \in \mathcal{J},\ g \in \mathcal{G},
\end{aligned}
\tag{34}
$$

where $\pi(g)$ is the predecessor group of group g. The estimations for the expectations are given by the means $\overline{p}^g, \overline{q}^g$, $\overline{h}^g$ and $\overline{f}^g$; the estimations for the variances are given by the sample unbiased variances $S^2_{q^g}$, $S^2_{h^g}$ and $S^2_{f^g}$ and the estimation of the covariance matrices $\hat{D}^{gg'}$ is given by $\hat{C}ov(q^g, h^g, f^g, q^{g'}, h^{g'}, f^{g'})$. The estimated covariance matrices $\hat{D}_{ij}^{gg'}$, $i, j \in \mathcal{J}$ are defined as,

$$\hat{D}_{ij}^{gg'} = \begin{pmatrix} S_{q_i^g q_j^g} & S_{q_i^g h_j^g} & S_{q_i^g f_j^g} & S_{q_i^g q_j^{g'}} & S_{q_i^g h_j^{g'}} & S_{q_i^g f_j^{g'}} \\ S_{h_i^g q_j^g} & S_{h_i^g h_j^g} & S_{h_i^g f_j^g} & S_{h_i^g q_j^{g'}} & S_{h_i^g h_j^{g'}} & S_{h_i^g f_j^{g'}} \\ S_{f_i^g q_j^g} & S_{f_i^g h_j^g} & S_{f_i^g f_j^g} & S_{f_i^g q_j^{g'}} & S_{f_i^g h_j^{g'}} & S_{f_i^g f_j^{g'}} \\ S_{q_i^{g'} q_j^g} & S_{q_i^{g'} h_j^g} & S_{q_i^{g'} f_j^g} & S_{q_i^{g'} q_j^{g'}} & S_{q_i^{g'} h_j^{g'}} & S_{q_i^{g'} f_j^{g'}} \\ S_{h_i^{g'} q_j^g} & S_{h_i^{g'} h_j^g} & S_{h_i^{g'} f_j^g} & S_{h_i^{g'} q_j^{g'}} & S_{h_i^{g'} h_j^{g'}} & S_{h_i^{g'} f_j^{g'}} \\ S_{f_i^{g'} q_j^g} & S_{f_i^{g'} h_j^g} & S_{f_i^{g'} f_j^g} & S_{f_i^{g'} q_j^{g'}} & S_{f_i^{g'} h_j^{g'}} & S_{f_i^{g'} f_j^{g'}} \end{pmatrix},$$

where $S_{q_i^g h_j^{g'}}$ is the sample-covariance between q_i^g and $h_j^{g'}$ for nodes $g, g' \in \mathcal{G}$, and so on.

This general multistage stochastic production planning problem has only quadratic terms for continuous variables. If quadratic terms are introduced between 0-1 variables and between 0-1 and continuous variables, reformulation techniques can be applied as explained in Section 2.3 and reference [10], to solve the problem efficiently.

4.2 Testbed results for time consistent EMV cost models

We report the results obtained in the computational experience while optimizing a testbed consisting of 15 medium and 15 large-scale problems with the structure described in Section 4.1 and five β values {0, 0.001, 0.01, 0.1, 1}. The thirty instances with $\beta = 1$ used for the computational experience are available in our repository `https://ehubox.ehu.eus/index.php/s/02Jhx3vYSXVQx7e`, see [6]. As far as we know, the first library for multistage stochastic mixed 0-1 quadratic problems, see [3] for two-stage stochastic integer programming and [21, 26] for stochastic linear programming.

The instances come from a realistic production planning problem, described in Section 4.1, whose dimensions are shown in Table 1 for medium-scale instances (Testbed 1) and in Table 2 for large-scale instances (Testbed 2). The headings are as follows: $Inst$, Instance code; T, number of stages; m, number of constraints; nx, number of 0-1 variables; nxc_i, number of common variables, i.e., number of 0-1 variables in the subset of stages up to the break stage $t^* = i$, for $i = 1, 2, 3$; ny, number of continuous variables; nel, number of non zero coefficients in the constraint matrix; $Qnel$, number of nonzero coefficients in the quadratic matrix; $|\Omega|$, number of scenarios; $|\mathcal{G}|$, number of nodes in the scenario tree; and $Tree$, scenario tree structure $A_1^{B_1} A_2^{B_2}$ or $A_1^{B_1} A_2^{B_2} A_3^{B_3}$, where A_i denotes the number of children at each node, B_i the number of stages where each node of the previous stage has A_i children $i = 1, 2, 3$, so the total number of stages is $T = 1+B_1+B_2$ or $T = 1+B_1+B_2+B_3$. In order to reduce the density of the quadratic matrix, we have considered D_{ij}^{gg} matrices; the decomposition scheme remains unaltered.

Tables 3 and 4 show the changes over time in the absolute and relative expected standard deviation costs versus the expected mean cost for the first eight medium-sized instances, where the optimal or best known solution is achieved, respectively. Table 3 details the mean-variance performance for the risk aversion parameter $\beta \geq 0$ in $\{0, 0.001, 0.01\}$ and Table 4 for β in $\{0, 0.1, 1\}$. The headings are as follows: $\hat{\mu}^\beta$, expected mean cost, i.e. value of linear part of the objective in (34), $\hat{\mu}^\beta = \sum_{g \in \mathcal{G}} w^g [\sum_{j \in \mathcal{J}} [\hat{\mu}_{p_j^g} X_j^g + \hat{\mu}_{q_j^g} Y_j^g + \hat{\mu}_{h_j^g} Z_j^g + \hat{\mu}_{f_j^g} F_j^g]]$; $\hat{\sigma}^\beta$, expected standard deviation of cost (absolute risk cost), i.e. the quadratic part of the objective function in (34) substracted from the risk parameter, $\hat{\sigma}^\beta = \sqrt{\sum_{g \in \mathcal{G}} w^g [\sum_{i,j \in \mathcal{J}} (Y_i^g, Z_i^g, F_i^g, Y_j^{g'}, Z_j^{g'}, F_j^{g'}) \hat{D}_{ij}^{gg'} (Y_i^g, Z_i^g, F_i^g, Y_j^{g'}, Z_j^{g'}, F_j^{g'})^t]}$; cv^β, the coefficient of variation for the cost (relative risk cost), i.e. the relative expected standard deviation with respect to the expected mean cost in percentage terms, $cv^\beta = \frac{\hat{\sigma}^\beta}{\hat{\mu}^\beta} 100$. Note that the objective function in (34) can be represented as $\hat{\mu}^\beta + \beta[\hat{\sigma}^\beta]^2$.

Table 1 Model dimensions for testbed 1

	T	m	nx	nxc_1	nxc_2	nxc_3	ny	nel	$Qnel$	$\|\Omega\|$	$\|\mathcal{G}\|$	$Tree$
QP1	5	4617	1110	10	60	210	2730	15662	38820	60	111	$5^1 3^1 2^2$
QP2	5	8679	2220	20	120	420	5460	33988	149640	60	111	$5^1 3^1 2^2$
QP3	5	13290	3510	30	150	630	8610	56853	351678	64	117	$4^2 2^2$
QP4	6	13992	3360	10	60	210	8280	47126	117600	180	336	$5^1 3^2 2^2$
QP5	5	17681	4680	40	200	840	11480	86802	620544	64	117	$4^2 2^2$
QP6	5	23370	6390	30	150	630	14850	101238	630174	144	213	$4^2 3^2$
QP7	6	26304	6720	20	120	420	16560	102316	453240	180	336	$5^1 3^2 2^2$
QP8	6	27933	7140	20	100	420	17580	114037	481440	192	357	$4^2 3^1 2^2$
QP9	5	31089	8520	40	200	840	19800	156071	1113216	144	213	$4^2 3^2$
QP10	6	33515	8050	10	130	490	19830	110490	281708	432	805	$12^1 3^2 2^2$
QP11	6	38280	10080	30	180	630	24840	161538	1011336	180	336	$5^1 3^2 2^2$
QP12	6	40650	10710	30	150	630	26370	173733	1074270	192	357	$4^2 3^1 2^2$
QP13	6	50928	13440	40	240	840	33120	248312	1784352	180	336	$5^1 3^2 2^2$
QP14	6	54081	14280	40	200	840	35160	265202	1895424	192	357	$4^2 3^1 2^2$
QP15	6	90131	21730	10	130	730	53190	297594	758540	1200	2173	$12^1 5^2 2^2$

Table 2 Model dimensions for testbed 2

	T	m	nx	nxc_1	nxc_2	nxc_3	ny	nel	$Qnel$	$\|\Omega\|$	$\|\mathcal{G}\|$	$Tree$
QP16	6	169397	43460	20	260	1460	106380	671243	2925120	1200	2173	$12^1 5^2 2^2$
QP17	5	194601	50920	40	1000	4840	122040	1113499	6708736	768	1273	$24^1 4^2 2^1$
QP18	6	227075	54850	10	130	1090	133830	739360	1912268	3072	5485	$12^1 8^2 2^2$
QP19	6	246490	65190	30	390	2190	159570	1057930	6527438	1200	2173	$12^1 5^2 2^2$
QP20	4	278935	77800	40	1000	16360	171960	1506990	10032640	1536	1945	$24^1 16^1 4^1$
QP21	5	294320	77800	40	1000	5800	185400	1590275	10234240	1200	1945	$24^1 5^2 2^1$
QP22	6	327929	86920	40	520	2920	212760	1605419	11518336	1200	2173	$12^1 5^2 2^2$
QP23	4	363490	105150	30	1470	12990	223290	1638250	10112678	3072	3505	$48^1 8^2$
QP24	5	367985	96500	20	500	4340	228060	1459700	6425760	3072	4825	$24^1 8^2 2^1$
QP25	4	413410	116670	30	1470	24510	257850	1802929	11343782	3072	3889	$48^1 16^1 4^1$
QP26	6	426725	109700	20	260	2180	267660	1665210	7376160	3072	5485	$12^1 8^2 2^2$
QP27	4	483485	140200	40	1960	17320	297720	2565210	17896960	3072	3505	$48^1 8^2$
QP28	4	496649	139220	20	980	16340	294780	2100077	8979360	6144	6961	$48^1 16^1 8^1$
QP29	5	583038	162280	40	1000	8680	363960	3093111	21005824	3072	4057	$24^1 8^1 4^2$
QP30	5	711845	193000	40	1000	8680	456120	3546650	25331200	3072	4825	$24^1 8^2 2^1$

Table 3 Risk-sensitive performance for $\beta \in \{0, 0.001, 0.01\}$

Inst.	$\beta = 0$			$\beta = 0.001$			$\beta = 0.01$		
	$\hat{\mu}^0$	$\hat{\sigma}^0$	cv^0	$\hat{\mu}^{0.001}$	$\hat{\sigma}^{0.001}$	$cv^{0.001}$	$\hat{\mu}^{0.01}$	$\hat{\sigma}^{0.01}$	$cv^{0.01}$
QP1	239563.2	2044.0	0.85	240245.2	1973.1	0.82	242116.8	1904.1	0.79
QP2	868069.8	1574.8	0.18	868400.0	1513.6	0.17	868542.7	1506.7	0.17
QP3	925046.8	6058.4	0.65	956176.2	5253.0	0.55	1022291.0	4241.3	0.41
QP4	276597.2	2743.7	0.99	277189.2	2664.8	0.96	279802.6	2602.1	0.93
QP5	1149530.1	9418.4	0.82	1190602.7	8718.3	0.73	1471930.6	5632.6	0.38
QP6	756935.1	5187.8	0.69	760836.6	5028.0	0.66	881012.8	2765.4	0.31
QP7	818333.6	3266.7	0.40	821502.1	3112.6	0.38	860785.3	2077.3	0.24
QP8	755795.4	2846.4	0.38	756133.9	2810.1	0.37	761777.0	2656.4	0.35

The results evidence a striking decrease in the absolute and relative risk costs, with an increase in the expected mean cost. Define the *mean.gap* and *sd.gap* as the relative difference between expected mean cost and expected standard deviation for $\beta > 0$ with respect to $\beta = 0$ in percentage terms, that is, $mean.gap = \frac{\hat{\mu}^{\beta} - \hat{\mu}^0}{\hat{\mu}^0} 100$ and $sd.gap = \frac{\hat{\sigma}^{\beta} - \hat{\sigma}^0}{\hat{\sigma}^0} 100$ for $\beta > 0$. Figure 3 shows the β changes over time in those gaps for instances of Tables 3 and 4. The figures for mean.gap versus sd.gap are represented, so the efficient frontier is illustrated (see [28]) in relative terms, i.e. with respect to values corresponding to $\beta = 0$. Position (0,0) represents the expected mean cost and the expected

Table 4 Risk-sensitive performance for $\beta \in \{0, 0.1, 1\}$

Inst.	$\beta = 0$			$\beta = 0.1$			$\beta = 1$		
	$\hat{\mu}^0$	$\hat{\sigma}^0$	cv^0	$\hat{\mu}^{0.1}$	$\hat{\sigma}^{0.1}$	$cv^{0.1}$	$\hat{\mu}^1$	$\hat{\sigma}^1$	cv^1
QP1	239563.2	2044.0	0.85	308016.2	1228.2	0.40	412498.2	945.7	0.23
QP2	868069.8	1574.8	0.18	895960.8	1348.8	0.15	947607.4	1243.5	0.13
QP3	925046.8	6058.4	0.65	1343195.6	2523.7	0.19	1563817.0	2306.6	0.15
QP4	276597.2	2743.7	0.99	410686.5	1439.0	0.35	565631.4	1127.0	0.20
QP5	1149530.1	9418.4	0.82	2001344.5	3252.6	0.16	2407408.5	2870.1	0.12
QP6	756935.1	5187.8	0.69	988184.7	2136.5	0.22	1264335.8	1851.2	0.15
QP7	818333.6	3266.7	0.40	921379.4	1363.2	0.15	964492.1	1314.3	0.14
QP8	755795.4	2846.4	0.38	848629.6	1692.6	0.20	1020233.0	1343.6	0.13

standard deviation for model (34) with $\beta = 0$, where the deviation is computed with the solution of the MIP model. Therefore, the regions that contain (0, 100) are inefficient regions while those that contain (-100,0) are infeasible regions. Note that the greater the risk parameter β is, the greater the increase in expected mean cost and the decrease in expected cost variability will be.

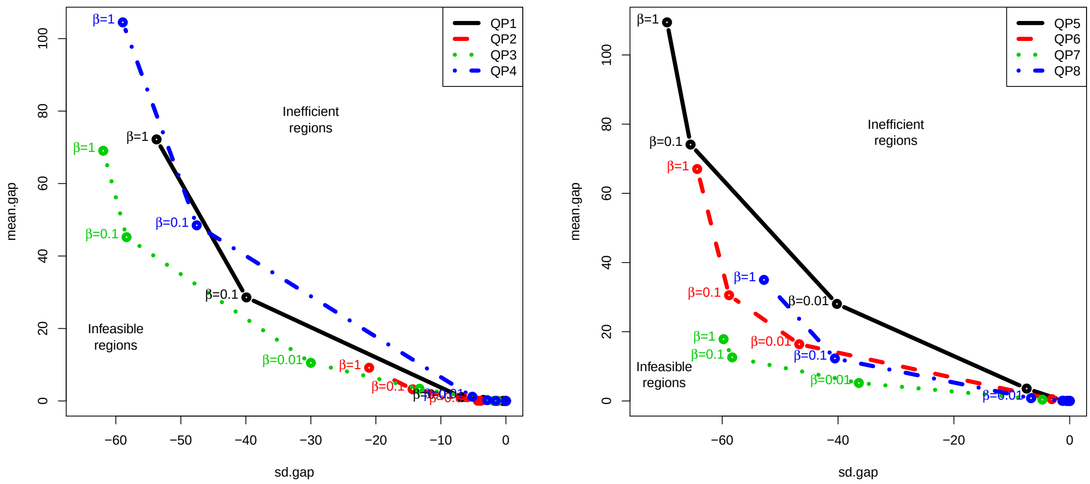

Fig. 3 Efficient frontier for problems QP1 to QP8.

4.3 Results of algorithm QH-BFC

Tables 5 to 8 show the performance of the QH-BFC algorithm compared to the plain use of CPLEX for the medium-sized instances of Testbed 1 whose dimensions are given in Table 1 and for $\beta \in \{0.001, 0.01, 0.1, 1\}$. The optimal gap used in CPLEX is 10^{-4}%. The headings are as follows: *Inst*, instance's code; z^{DEM}_{CPLEX}, solution value of the original model (34) using CPLEX; *ti* time in seconds after which the CPLEX incumbent solution is found; t, elapsed time in seconds; *OG*, optimality gap in percentage terms shown by CPLEX; z^{DEM}, solution value of the original model (34) using QH-BFC; t, elapsed time in seconds; *GG*, goodness gap in percentage terms shown by QH-BFC, i.e. relative difference in percentage terms between the QH-BFC and CPLEX solution values, $GG = \frac{z^{DEM} - z^{DEM}_{CPLEX}}{z^{DEM}_{CPLEX}} \cdot 100$. Note that the elapsed time, t, for QH-BFC algorithm includes the time required to generate the cluster quadratic submodels.

Tables 5 to 8 report the optimal solution for medium-sized instances from QP1 to QP15 for several break stages t^*. They are solved using the QH-BFC algorithm with small goodness gaps and very competitive CPU times compared to the plain use of CPLEX. Table 5 illustrates the performance of the algorithm for the cases of maximum risk ($\beta = 0.001$) and Table 8 for those with low level of risk ($\beta = 1$). The intermediate values are shown in Tables 6 and 7.

Table 5 Performance of QH-BFC for Testbed 1 and $\beta = 0.001$

Inst.	Plain CPLEX				QH-BFC, $t^* = 1$			QH-BFC, $t^* = 2$			QH-BFC, $t^* = 3$		
Inst	z^{DEM}_{CPLEX}	ti	t	OG	z^{DEM}	t	GG	z^{DEM}	t	GG	z^{DEM}	t	GG
QP1	243711.5	2	2	*	244933.3	1	0.50	246435.9	1	1.12	256977.5	2	5.44
QP2	870534.1	1	1	*	873871.3	1	0.38	874161.2	1	0.42	882525.7	3	1.38
QP3	960542.8	3	4	*	964375.5	4	0.40	968653.9	3	0.84	978937.5	5	1.92
QP4	283957.8	36	37	*	283988.6	3	0.01	285865.7	2	0.67	285815.9	4	0.65
QP5	1236776.0	6	6	*	1237698.6	5	0.07	1240422.9	5	0.29	1241078.6	6	0.35
QP6	783687.4	17	18	*	784144.2	6	0.06	785988.8	6	0.29	790000.0	7	0.81
QP7	828992.7	2	7	*	829751.2	3	0.09	829751.2	3	0.09	831519.9	5	0.30
QP8	763824.8	156	157	*	764565.7	6	0.10	772661.0	5	1.16	776181.6	4	1.62
QP9	1372164.9	6	7	*	1372164.9	7	*	1373279.5	6	0.08	1375027.2	8	0.21
QP10	352263.4	8276	8277	*	352502.2	6	0.07	352867.1	4	0.17	353693.0	7	0.41
QP11	939487.8	3181	3334	*	941033.9	13	0.16	944523.6	11	0.54	947608.6	11	0.86
QP12	956831.2	654	664	*	963648.0	17	0.71	970507.1	15	1.43	972243.7	14	1.61
QP13	1761050.4	10	11	*	1761050.4	14	*	1762660.7	12	0.09	1764831.1	12	0.21
QP14	1235600.7	12175	13145	*	1235825.5	65	0.02	1243120.3	30	0.61	1245216.1	30	0.78
QP15	444067.9	245	262	*	444067.9	11	*	446802.8	9	0.62	448982.2	22	1.11

*: optimality gap achieved ($\leq 0.0001\%$)

Table 6 Performance of QH-BFC for Testbed 1 and $\beta = 0.01$

Inst.	Plain CPLEX				QH-BFC, $t^* = 1$			QH-BFC, $t^* = 2$			QH-BFC, $t^* = 3$		
	z^{DEM}_{CPLEX}	ti	t	OG	z^{DEM}	t	GG	z^{DEM}	t	GG	z^{DEM}	t	GG
QP1	278374.5	10	10	*	279867.6	1	0.54	283079.4	2	1.69	288912.8	2	3.79
QP2	891245.7	2	2	*	893977.4	2	0.31	896465.2	1	0.59	901354.4	3	1.13
QP3	1202179.0	171	212	*	1203734.6	10	0.13	1207429.1	6	0.44	1215708.3	8	1.13
QP4	347513.5	3220	5792	*	347513.5	5	*	348236.5	3	0.21	348683.0	4	0.34
QP5	1789197.0	4203	4257	*	1789197.0	18	*	1789327.1	10	0.01	1791895.9	11	0.15
QP6	957485.9	1373	1561	*	957485.9	12	*	958188.5	6	0.07	960010.9	7	0.26
QP7	903937.6	441	851	*	904675.1	8	0.08	904675.1	6	0.08	905383.7	6	0.16
QP8	832342.2	4328	5190	*	832874.2	18	0.06	839731.4	8	0.89	842564.4	8	1.23
QP9	1804267.1	19458	21600	1.29	-	21600	-	1805497.3	25	0.07	1805212.2	22	0.05
QP10	397826.5	20233	21600	0.76	397803.3	15	-0.01	398276.3	7	0.11	398536.9	9	0.18
QP11	1178244.9	19857	21600	3.17	1179389.8	16178	0.10	1182028.8	53	0.32	1185153.7	27	0.59
QP12	1292080.9	19978	21600	2.69	-	21600	-	1298811.5	66	0.52	1299030.0	35	0.54
QP13	2173599.4	19536	21600	1.18	-	21600	-	2175007.6	98	0.06	2176892.2	33	0.15
QP14	1956025.3	21195	21600	1.94	-	21600	-	1955565.8	1177	-0.02	1956484.4	166	0.02
QP15	482827.2	19960	21600	0.24	482795.8	27	-0.01	484595.5	17	0.37	486863.6	28	0.84

*: optimality gap achieved ($\leq 0.0001\%$)
-: z^{DEM} not available, exceeded time limit (21600 secs.)

Table 7 Performance of QH-BFC for Testbed 1 and $\beta = 0.1$

	Plain CPLEX				QH-BFC, $t^* = 1$			QH-BFC, $t^* = 2$			QH-BFC, $t^* = 3$		
Inst	z^{DEM}_{CPLEX}	*ti*	*t*	*OG*	z^{DEM}	*t*	*GG*	z^{DEM}	*t*	*GG*	z^{DEM}	*t*	*GG*
QP1	458853.6	15	21600	0.64	458853.6	4	0.00	459085.7	2	0.05	466375.5	2	1.64
QP2	1077894.4	3	3	*	1079521.8	3	0.15	1079525.1	2	0.15	1086438.6	3	0.79
QP3	1980108.3	310	21600	0.61	1980108.3	54	0.00	1984200.3	8	0.21	1985167.3	10	0.26
QP4	617744.9	20021	21600	2.67	617726.0	6734	-0.00	617734.5	10	-0.00	617790.0	6	0.01
QP5	3059269.8	5059	21600	2.18	3059024.1	84	-0.01	3059483.2	18	0.01	3060419.9	16	0.04
QP6	1444666.3	1796	5364	*	1444683.2	36	0.00	1445920.8	9	0.09	1446203.5	10	0.11
QP7	1107213.3	317	320	*	1107728.7	15	0.05	1107728.7	8	0.05	1108315.5	8	0.10
QP8	1135116.8	20111	21600	3.14	1134158.3	18674	-0.08	1139056.9	21	0.35	1137562.8	12	0.22
QP9	2697515.1	19463	21600	2.92	-	21600	-	2697028.1	39	-0.02	2696819.5	22	-0.03
QP10	592853.6	937	21600	2.78	592846.7	36	-0.00	593018.9	9	0.03	593058.1	8	0.03
QP11	1941219.6	21118	21600	1.36	-	21600	-	1941959.5	86	0.04	1941959.5	29	0.04
QP12	2101147.6	19802	21600	3.79	2096506.0	21600	-0.22	2096594.7	88	-0.22	2097886.9	39	-0.16
QP13	3385430.3	15325	21600	3.55	-	21600	-	3384226.3	233	-0.04	3384409.0	68	-0.03
QP14	3332394.3	19944	21600	1.78	-	21600	-	3332300.0	261	-0.00	3332644.2	147	0.01
QP15	636088.7	19635	21600	1.73	636080.5	195	-0.00	636771.5	25	0.11	637295.4	28	0.19

*: optimality gap achieved ($\leq 0.0001\%$)
-: z^{DEM} not available, exceeded time limit (21600 secs.)

Table 8 Performance of QH-BFC for Testbed 1 and $\beta = 1$

	Plain CPLEX				QH-BFC, $t^* = 1$			QH-BFC, $t^* = 2$			QH-BFC, $t^* = 3$		
Inst	z^{DEM}_{CPLEX}	*ti*	*t*	*OG*	z^{DEM}	*t*	*GG*	z^{DEM}	*t*	*GG*	z^{DEM}	*t*	*GG*
QP1	1306804.3	8945	21600	0.61	1306835.7	5	0.00	1306835.8	2	0.00	1311180.0	2	0.33
QP2	2493837.8	9	279	*	2493949.8	3	0.00	2494785.0	2	0.04	2495581.0	2	0.07
QP3	6884242.4	19493	21600	0.38	6884242.4	140	0.00	6886279.5	13	0.03	6887225.5	13	0.04
QP4	1835690.3	19642	21600	4.40	1835530.0	269	-0.01	1835530.0	6	-0.01	1835530.0	4	-0.01
QP5	10644823.9	19783	21600	0.81	10644732.3	1640	-0.00	10644866.8	33	0.00	10647284.3	23	0.02
QP6	4691252.6	19587	21600	1.59	-	21600	-	4695608.4	72	0.09	4702455.3	27	0.24
QP7	2691885.5	20804	21600	0.39	2692819.8	296	0.03	2692974.9	23	0.04	2694350.4	14	0.09
QP8	2825555.3	19729	21600	4.07	2825033.9	21600	-0.02	2825061.2	27	-0.02	2826141.4	13	0.02
QP9	8404206.1	12	21600	1.84	8404206.1	380	0.00	8404206.1	19	0.00	8404206.1	15	0.00
QP10	1537521.5	19712	21600	0.68	1537522.9	72	0.00	1540518.2	13	0.19	1540750.6	10	0.21
QP11	6546510.2	20083	21600	3.51	-	21600	-	6548148.0	2732	0.03	6549728.7	58	0.05
QP12	7033987.9	19498	21600	6.58	-	21600	-	7035452.3	95	0.02	7036023.3	42	0.03
QP13	11582579.6	20202	21600	4.38	11565213.1	1431	-0.15	11565213.1	104	-0.15	11568139.1	42	-0.12
QP14	11707481.0	21418	21600	2.61	-	21600	-	11707469.5	8217	-0.00	11706324.4	305	-0.01
QP15	1669934.9	21185	21600	4.99	-	21600	-	1668792.2	41	-0.07	1669725.9	32	-0.01

*: optimality gap achieved ($\leq 0.0001\%$)
-: z^{DEM} not available, exceeded time limit (21600 secs.)

The results show that the larger the risk parameter is, the harder it is to solve the problems by the plain use of CPLEX. The QH-BFC performance in terms of the break stage reports the best quality results for $t^* = 1$ and $t^* = 2$, while the shortest elapsed times are reported for $t^* = 2$ and $t^* = 3$. Note that a negative GG indicates that better solutions are obtained by QH-BFC than with the plain use of CPLEX: three instances for $\beta = 0.001$, nine for $\beta = 0.1$ and six for $\beta = 1$. Moreover, the QH-BFC results are better than or equal to those obtained with the plain use of CPLEX in more than half of the medium-sized instances.

Tables 9 to 12, report the performance of QH-BFC versus plain use of CPLEX for the large-sized instances, QP16 to QP30. On the one hand, the tables show the execution of the QH-BFC algorithm with the break stage where the smallest goodness gaps are obtained and on the other hand the results of using QH-BFC with McCormick relaxation for $m^* = 1$ and the break stage, where the shortest CPU times are achieved.

Table 9 Performance of QH-BFC for Testbed 2 and $\beta = 0.001$

	Plain CPLEX				QH-BFC, *min GG%*				QH-BFC McCormick, *min t*			
Inst	z^{DEM}_{CPLEX}	*ti*	*t*	*OG*	t^*	z^{DEM}	*t*	*GG*	t^*	z^{DEM}	t	*GG*
QP16	732860.5	21413	21600	0.11	1	736376.2	370	0.48	2	738661.2	49	0.79
QP17	1594367.5	8224	8227	*	1	1596179.8	55	0.11	2	1597852.6	46	0.22
QP18	451406.6	19914	21600	0.03	1	451406.6	98	0.00	2	454895.4	37	0.77
QP19	1178053.2	21355	21600	0.02	1	1178055.8	103	0.00	2	1180070.0	64	0.17
QP20	1136767.5	799	804	*	1	1138569.6	65	0.16	1	1138600.8	65	0.16
QP21	1239717.4	21412	21600	0.05	1	1240706.3	114	0.08	2	1242612.5	83	0.23
QP22	1492045.2	19812	21600	0.14	1	1496834.2	885	0.32	2	1500170.2	132	0.54
QP23	900724.5	81	91	*	1	901081.6	56	0.04	1	901081.6	56	0.04
QP24	759010.3	20835	21600	0.05	1	758989.4	118	-0.00	2	762037.8	82	0.40
QP25	876224.0	16967	21600	0.00	1	877170.7	93	0.11	1	877170.7	93	0.11
QP26	873485.9	19890	21600	0.20	1	873844.9	3474	0.04	2	879891.0	125	0.73
QP27	1407885.8	190	228	*	1	1407885.8	97	*	1	1407885.8	96	*
QP28	604057.5	97	103	*	1	604057.5	89	*	1	604057.5	87	*
QP29	1559059.6	19494	21600	0.02	1	1559592.2	216	0.03	2	1563059.3	177	0.26
QP30	1476134.8	20191	21600	0.08	1	1477437.3	539	0.09	2	1480215.3	290	0.28

* optimality gap achieved ($\leq 0.0001\%$)

Table 10 Performance of QH-BFC for Testbed 2 and $\beta = 0.01$

	Plain CPLEX				QH-BFC, *min GG%*				QH-BFC McCormick, *min t*			
Inst	z^{DEM}_{CPLEX}	*ti*	*t*	*OG*	t^*	z^{DEM}	*t*	*GG*	t^*	z^{DEM}	*t*	*GG*
QP16	833215.7	21600	21600	1.06	2	833389.5	149	0.02	3	834265.5	97	0.13
QP17	1911416.2	3672	21600	2.17	2	1911668.5	229	0.01	3	1912790.6	130	0.07
QP18	467704.7	20753	21600	0.05	1	467691.7	202	-0.00	2	469080.2	50	0.29
QP19	1369928.1	21600	21600	1.46	2	1371067.8	368	0.08	3	1372103.8	131	0.16
QP20	1435153.9	4141	21600	0.17	1	1436752.8	277	0.11	2	1445442.2	133	0.72
QP21	1746823.7	14191	21600	2.70	2	1747178.2	1356	0.02	3	1748148.2	277	0.08
QP22	2139856.8	19203	21600	4.05	3	2140560.0	892	0.03	3	2140931.3	724	0.05
QP23	1046962.1	20124	21600	1.50	2	1048263.6	128	0.12	2	1048282.7	116	0.13
QP24	833579.8	21600	21600	1.66	2	833682.4	440	0.01	2	833674.4	155	0.01
QP25	1053973.8	20079	21600	0.31	1	1054526.6	678	0.05	2	1054724.1	217	0.07
QP26	935093.9	20097	21600	0.37	3	937959.4	547	0.31	2	940117.1	355	0.54
QP27	1674943.1	4142	21600	0.40	1	1674929.7	296	-0.00	2	1675393.8	175	0.03
QP28	622249.3	122	171	*	1	622249.3	96	*	1	622249.3	92	*
QP29	1968212.8	18099	21600	0.51	2	1969434.3	653	0.06	2	1980872.1	419	0.64
QP30	1956735.8	21352	21600	1.98	2	1943471.5	1893	-0.68	2	1943506.9	1832	-0.68

* optimality gap achieved ($\leq 0.0001\%$)

Table 11 Performance of QH-BFC for Testbed 2 and $\beta = 0.1$

	Plain CPLEX				QH-BFC, *min GG%*				QH-BFC McCormick, *min t*			
Inst	z^{DEM}_{CPLEX}	*ti*	*t*	*OG*	t^*	z^{DEM}	*t*	*GG*	t^*	z^{DEM}	*t*	*GG*
QP16	1161160.3	21600	21600	1.65	2	1159976.7	553	-0.10	3	1161536.0	155	0.03
QP17	2784280.0	424	21600	1.64	1	2783971.5	488	-0.01	2	2799706.3	182	0.55
QP18	605070.1	21600	21600	1.51	2	605039.1	145	-0.01	2	605446.0	82	0.06
QP19	1905205.4	21600	21600	0.84	2	1902669.1	540	-0.13	3	1905281.8	234	0.00
QP20	2205766.1	21600	21600	0.88	1	2206286.5	6282	0.02	2	2206872.7	144	0.05
QP21	2740798.5	21600	21600	2.58	2	2739933.2	844	-0.03	3	2749412.9	403	0.31
QP22	3232869.6	21600	21600	3.07	2	3232388.4	3194	-0.01	3	3236984.6	790	0.13
QP23	1400862.5	9178	21600	0.91	1	1402583.2	251	0.12	2	1403072.1	147	0.16
QP24	1057121.9	7933	21600	1.26	2	1057492.1	330	0.04	2	1057495.6	201	0.04
QP25	1586620.0	5196	21600	0.62	1	1586609.7	534	-0.00	2	1599283.2	268	0.80
QP26	1265315.7	21600	21600	2.26	2	1264527.6	1561	-0.06	3	1265113.0	488	-0.02
QP27	2347642.7	20258	21600	1.24	1	2347474.5	588	-0.01	2	2348187.7	239	0.02
QP28	783631.3	21003	21600	0.07	1	783616.1	183	-0.00	1	783622.9	148	-0.00
QP29	2924122.3	6288	21600	2.10	2	2924431.4	1103	0.01	2	2925183.7	973	0.04
QP30	2999935.3	19228	21600	3.93	2	2982077.5	2837	-0.60	2	2983062.9	2507	-0.56

* optimality gap achieved ($\leq 0.0001\%$)

Table 12 Performance of QH-BFC for Testbed 2 and $\beta = 1$

	Plain CPLEX				QH-BFC, *min GG%*				QH-BFC McCormick, *min t*			
Inst	z^{DEM}_{CPLEX}	*ti*	*t*	*OG*	t^*	z^{DEM}	*t*	*GG*	t^*	z^{DEM}	*t*	*GG*
QP16	2889528.0	21600	21600	6.35	3	2890039.2	272	0.02	3	2903534.0	263	0.48
QP17	9191798.3	10536	21600	1.76	1	9190512.8	1024	-0.01	3	9450182.2	185	2.81
QP18	1603533.0	21600	21600	1.86	2	1602967.7	291	-0.04	3	1611100.2	113	0.47
QP19	5966757.4	14842	21600	2.92	3	5974940.9	594	0.14	3	6008316.7	433	0.70
QP20	7766676.7	21138	21600	1.64	2	7769849.1	282	0.04	2	7775099.5	170	0.11
QP21	9404084.9	17784	21600	3.26	2	9402986.0	976	-0.01	3	9496599.7	444	0.98
QP22	10328287.7	21522	21600	4.46	3	10344398.8	1152	0.16	3	10385346.1	1131	0.55
QP23	4156502.5	20976	21600	1.40	1	4189383.5	453	0.79	2	4202465.4	125	1.11
QP24	2526841.1	6126	21600	6.18	2	2528825.4	608	0.08	2	2532748.7	394	0.23
QP25	5317381.0	5649	21600	0.92	2	5323276.1	559	0.11	2	5457308.8	227	2.63
QP26	3165474.8	16634	21600	4.07	2	3158526.0	5767	-0.22	3	3164071.8	490	-0.04
QP27	7337863.1	21202	21600	1.78	1	7337783.1	2329	-0.00	2	7347437.2	207	0.13
QP28	2009324.3	20717	21600	4.14	2	2009553.2	479	0.01	2	2009806.6	233	0.02
QP29	9258501.8	11812	21600	2.35	2	9317154.0	1019	0.63	2	9325607.5	753	0.72
QP30	9872316.5	19471	21600	5.27	2	9851139.4	4447	-0.21	2	9856311.2	2201	-0.16

* optimality gap achieved ($\leq 0.0001\%$)

The QH-BFC algorithm finds better solutions than CPLEX in a third of the instances, as evidenced by the negative goodness gaps with short CPU times. For low level of risk, $\beta = 1$, which has the greatest weight in the quadratic term, the quality of QH-BFC solutions is better than that of the CPLEX solution in the instances QP4, QP5, QP8, QP13, QP14 and QP15. For risk levels, $\beta = 0.001$ and $\beta = 0.01$, the algorithm drastically reduces CPU times in the instances QP16 to QP30 and obtains solutions with very small goodness gaps. For $\beta = 0.1$ and the instances QP16 to QP19, QP21, QP22, QP25 to QP28 the algorithm drastically reduces CPU times and obtains better solutions than CPLEX. For $\beta = 1$ and the instances QP17, QP18, QP21, QP26, QP27 and QP30 the algorithm drastically reduces CPU times and obtains better solutions than CPLEX.

5 Conclusions and future work

This paper introduces the Expected Mean-Variance model, which is a multistage time consistent ECRM. It also presents a Quadratic mathHeuristic Branch-and-Fix Coordination algorithmic framework, named *QH-BFC*, for solving multistage quadratic mixed 0-1 problems under uncertainty. This technique is very suitable for solving multistage stochastic models with Expected Conditional Risk Measures. We consider stochasticity from two perspectives. First, we use a scenario cluster analysis approach to introduce uncertainty into the parameters in the objective function, rhs and constraint matrix coefficients. Second, we consider the distribution functions of the coefficients of the objective function along each scenario, using probability and estimation techniques, to introduce expected risk into the minimization objective. We believe that real-life uncertainty is very frequently represented via a scenario analysis approach. Also, scenario cluster analysis is crucial for solving medium and large scale quadratic problems, since it allows us to decompose such problems. The broad computational experience reported assesses the quality of the matheuristic solution for medium and large-scale multistage stochastic quadratic mixed 0-1 problems. The outstanding solution quality and computation times found in the new approach show competitive results when compared with the-state-of-the-art commercial optimizer CPLEX.

As a future line of work we are considering incorporating and comparing other expected conditional risk measures for multistage stochastic quadratic problems, and would like to extend the theory and QH-BFC algorithmic framework for solving problems subject to linear and quadratic constraints.

Acknowledgements The authors thank for technical and human support provided by IZO-SGI SGIker of UPV/EHU and European funding (ERDF and ESF).

References

1. Adams, W.P., Forrester, R.J., Glover, F.W.: Comparisons and enhancement strategies for linearizing mixed 0-1 quadratic programs. Discrete Optimization **1**(2), 99–120 (2004). DOI 10.1016/j.disopt.2004.03.006
2. Ahmed, S.: Convexity and decomposition of mean-risk stochastic programs. Mathematical Programming **106**(3), 433–446 (2006). DOI 10.1007/s10107-005-0638-8
3. Ahmed, S., Garcia, R., Kong, N., Ntaimo, L., Parija, G., Qiu, F., Sen, S.: SIPLIB: A Stochastic Integer Programming Test Problem Library (2015). URL `http://www.isye.gatech.edu/~sahmed/siplib`
4. Aldasoro, U., Escudero, L.F., Merino, M., Pérez, G.: An algorithmic framework for solving large-scale multistage stochastic mixed 0-1 problems with nonsymmetric scenario trees. Part II: Parallelization. Computers & Operations Research **40**, 2950–2960 (2013). DOI 10.1016/j.cor.2013.06.015
5. Aldasoro, U., Escudero, L.F., Merino, M., Pérez, G.: A parallel Branch-and-Fix Coordination based matheuristic algorithm for solving large sized multistage stochastic mixed 0-1 problems. European Journal of Operational Research **258**(2), 590–606 (2017). DOI 10.1016/j.ejor.2016.08.072
6. Aldasoro, U., Merino, M., Pérez, G.: SMIQLib: Dataset for Stochastic Mixed Integer Quadratic Optimization. Website (2017). `https://ehubox.ehu.eus/index.php/s/02Jhx3vYSXVQx7e`
7. Alonso-Ayuso, A., Escudero, L.F., Garín, M.A., Ortuño, M.T., Pérez, G.: On the product selection and plant dimensioning problem under uncertainty. Omega **33**(4), 307–318 (2005). DOI 10.1016/j.omega.2004.05.001
8. ARINA: Computational cluster from izo-sgi, sgiker (upv/ehu) (2017). URL `http://www.ehu.eus/sgi/recursos/cluster-arina`
9. Birge, J.R., Louveaux, F.: Introduction to Stochastic Programming, 2nd edn. Springer (2011). DOI 10.1007/978-1-4614-0237-4
10. Bliek, C., Bonami, P., Lodi, A.: Solving Mixed-Integer Quadratic Programming problems with IBM-CPLEX : a progress report. In: Proceedings of the Twenty-Sixth RAMP Symposium, pp. 171–180 (2014)
11. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press (Seventh printing with corrections 2009)
12. Cesarone, F., Scozzari, A., Tardella, F.: A new method for mean-variance portfolio optimization with cardinality constraints. Annals of Operations Research **205**(1), 213–234 (2013). DOI 10.1007/s10479-012-1165-7
13. Conejo, A.J., Nogales, F.J., Arroyo, J.M., García-Bertrand, R.: Risk-constrained self-scheduling of a thermal power producer. IEEE Transactions on Power Systems **19**(3), 1569–1574 (2004). DOI 10.1109/TPWRS.2004.831652
14. Cristobal, M.P., Escudero, L.F., Monge, J.F.: On stochastic dynamic programming for solving large-scale planning problems under uncertainty. Computers & Operations Research **36**, 2418–2428 (2009). DOI 10.1016/j.cor.2008.09.009

15. Dentcheva, D., Ruszczyński, A.: Optimization with multivariate stochastic dominance constraints. Mathematical Programming **117**(1), 111–127 (2009). DOI 10.1007/s10107-007-0165-x
16. Dentcheva, D., Ruszczyński, A.: Robust stochastic dominance and its application to risk-averse optimization. Mathematical Programming **123**(1), 85 (2009). DOI 10.1007/s10107-009-0321-6
17. Escudero, L., Garín, M., Merino, M., Pérez, G.: A general algorithm for solving two-stage stochastic mixed 0-1 first-stage problems. Computers & Operations Research **36**(9), 2590 – 2600 (2009). DOI 10.1016/j.cor.2008.11.011
18. Escudero, L.F., Garín, M.A., Merino, M., Pérez, G.: BFC-MSMIP: an exact branch-and-fix coordination approach for solving multistage stochastic mixed 0-1 problems. TOP **17**, 96–122 (2009). DOI 10.1007/s11750-009-0083-6
19. Escudero, L.F., Garín, M.A., Merino, M., Pérez, G.: On BFC-MSMIP strategies for scenario cluster partitioning, and twin node family branching selection and bounding for multistage stochastic mixed integer programming. Computers & Operations Research **37**(4), 738–753 (2010). DOI 10.1016/j.cor.2009.06.023
20. Escudero, L.F., Garín, M.A., Merino, M., Pérez, G.: An algorithmic framework for solving large-scale multistage stochastic mixed 0-1 problems with nonsymmetric scenario trees. Computers & Operations Research **39**, 1133–1144 (2012). DOI 10.1016/j.cor.2011.06.021
21. Felt, A.: Test-Problem Collection for Stochastic Linear Programming (2003). URL `https://www4.uwsp.edu/math/afelt/slptestset/download.html`
22. Fortet, R.: L'algebre de boole et ses applications en recherche operationnelle. Revue française d'informatique et de recherche opérationnelle **11**(2), 111–118 (1960). DOI 10.1007/BF03006558
23. Furini, F., Traversi, E., Belotti, P., Frangioni, A., Gleixner, A., Gould, N., Liberti, L., Lodi, A., Misener, R., Mittelmann, H., Sahinidis, N.V., Vigerske, S., Wiegele, A.: QPLIB: A Library of Quadratic Programming Instances. URL `http://qplib.zib.de/`
24. Gantmacher, F.R.: The theory of matrices. Chelsea Publishing Company, New York (1959)
25. Glover, F.: Improved linear integer programming formulations of nonlinear integer problems. Management Science **22**(4), 455–460 (1975). DOI 10.1287/mnsc.22.4.455
26. Holmes, D.: A (PO)rtable (S)tochastic programming (T)est (S)et (POSTS) (2017). URL `http://users.iems.northwestern.edu/~jrbirge/html/dholmes/post.html`
27. IBM ILOG: CPLEX v12.6.3 (2017). URL `http://www.ilog.com/products/cplex`
28. Kall, P., Wallace, S.W.: Stochastic Programming. Wiley (1994)
29. Kolodziej, S., Castro, P.M., Grossmann, I.E.: Global optimization of bilinear programs with a multiparametric disaggregation technique. Journal of Global Optimization **57**(4), 1039–1063 (2013). DOI 10.1007/s10898-012-0022-1
30. Louveaux, F.V.: A solution method for multistage stochastic programs with recourse with application to an energy investment problem. Operations Research **29**(4), 889–902 (1980). DOI 10.1287/opre.28.4.889
31. McCormick, G.P.: Computability of global solutions to factorable nonconvex programs: Part I - Convex underestimating problems. Mathematical Programming **10**(1), 147–175 (1976). DOI 10.1007/BF01580665
32. de Mello, T.H., Pagnoncelli, B.K.: Risk aversion in multistage stochastic programming: A modeling and algorithmic perspective. European Journal of Operational Research **249**(1), 188 – 199 (2016). DOI http://dx.doi.org/10.1016/j.ejor.2015.05.048
33. Mijangos, E.: An algorithm for two-stage stochastic mixed-integer nonlinear convex problems. Annals of Operations Research **235**(1), 581–598 (2015). DOI 10.1007/s10479-015-1899-0
34. Mula, J., Poler, R., García-Sabater, J.P., Lario, F.C.: Models for production planning under uncertainty: A review. International Journal of Production Economics **103**(1), 271–285 (2006). DOI 10.1016/j.ijpe.2005.09.001
35. Neise, F.: Risk Management in Stochastic Integer Programming. Vieweg+Teubner Verlag (2008)
36. Osorio, M.A., Gulpinar, N., Rustem, B.: A mixed integer programming model for multistage mean-variance post-tax optimization. European Journal of Operational Research **185**(2), 451–480 (2008). DOI 10.1016/j.ejor.2006.09.105
37. Osorio, M.A., Gülpınar, N., Rustem, B.: A general framework for multistage mean-variance post-tax optimization. Annals of Operations Research **157**(1), 3–23 (2008). DOI 10.1007/s10479-007-0255-4
38. Pflug, G.C., Römisch, W.: Modeling, measuring and managing risk. World Scientific Publishing Co. Inc (2007)
39. Pochet, Y., Wolsey, L.A.: Production Planning by Mixed Integer Programming. Springer (2006). DOI 10.1007/0-387-33477-7
40. Shapiro, A.: On a time consistency concept in risk averse multistage stochastic programming. Operations Research Letters **37**(3), 143 – 147 (2009). DOI 10.1016/j.orl.2009.02.005
41. Siddiqui, S., Gabriel, S.A., Azarm, S.: Solving mixed-integer robust optimization problems with interval uncertainty using Benders decomposition. Journal of the Operational Research Society **66**(4), 664–673 (2015). DOI 10.1057/jors.2014.41
42. Sun, J., Liao, L.Z., Rodrigues, B.: Quadratic two-stage stochastic optimization with coherent measures of risk. Mathematical Programming pp. 1–15 (2017). DOI 10.1007/s10107-017-1131-x
43. Wets, R.J.B.: Stochastic Programs with Fixed Recourse: The Equivalent Deterministic Program. SIAM Review **16**(3), 309–339 (1974). DOI 10.1007/s10107-012-0621-0
44. Wets, R.J.B.: On the Relation between Stochastic and Deterministic Optimization. In: Control Theory, Numerical Methods and Computer Systems Modelling, pp. 350–361. Springer (1975). DOI 10.1007/978-3-642-46317-4_26
45. Wu, B., Sun, X., Li, D., Zheng, X.: Tight MIQP Reformulations for Semi-Continuous Quadratic Programming: Lift-and-Convexification Approach. arXiv:1507.05708v1 [math.OC] (2015)