Generator maintenance scheduling based on expected energy production

J. Eygelaar & J.H. van Vuuren
Stellenbosch Unit for Operations Research in Engineering, Department of Industrial Engineering, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa.

Abstract

Unexpected breakdowns of power generating units (PGUs) may significantly influence the capability of power utilities to satisfy energy demand. Due to low system capacity and high energy demand, planned maintenance of PGUs is often neglected despite the fact that unexpected failures are, in most cases, more expensive to repair than taking planned preventative maintenance action. The typical scheduling criteria pursued in the design of PGU maintenance schedules do not, however, take these difficulties into account. A new PGU maintenance scheduling criterion is therefore proposed in this paper. The scheduling objective proposed seeks to maximise the expected energy production within a scheduling window. The timings of power generating failures are determined by methods from reliability theory and are modelled by the incorporation of random variables in the model, facilitating the calculation of an expected failure time for each PGU. The effectiveness of the newly proposed scheduling objective is analysed by applying it to two well-known generating maintenance scheduling test systems from the literature. Since the scheduling objective is nonlinear, a piecewise linearisation component is incorporated in the solution approach.

Keywords: Power systems, preventative generator maintenance scheduling, reliability, expected energy, unit failure.

1. Introduction

In the well-known generator maintenance scheduling (GMS) problem, the aim is to develop maintenance schedules for the power generating units (PGUs) of a power system in such a manner that the system demand is satisfied effectively and efficiently [1]. In developing countries, this is a difficult task and the maintenance of these PGUs are often neglected due to small system capacities, where maintenance cannot always be performed during high demand periods. This causes PGUs to operate under high strain, often resulting in failures of these PGUs.

South Africa, which may be cited as an extreme case in point, has recently found it very difficult to construct maintenance schedules without exposing the integrity of the power grid, as exemplified by the load shedding programme implemented by the South African national power utility in 2008 and again in 2015. This programme involved rolling blackouts for certain areas of the country during certain times of the day to ensure that the system demand is satisfiable. The programme did not only reduce the energy demand of the country, but also ensured that PGU breakdowns were avoided and that preventative maintenance could be performed on high risk PGUs [2]. The current state of the power system has recently improved, but this situation could have been avoided altogether by taking into account the failure rates and expected failure times of the PGUs in the system when scheduling planned preventative maintenance procedures.

Most of the recent developments in GMS model formulations have, however, been focused on minimising production cost or maximising the profits of power utilities. The novel GMS criterion proposed in this paper is rather aimed...
at maximising the expected energy production over the
scheduling window. The occurrences of PGU failures are
modelled as random variables, taking into account three
possible cases of failures. The first case is where a PGU
failure is observed before planned maintenance is sched-
uled to be performed on that PGU. The second case is
where a PGU failure is observed after planned mainte-
nance has already been performed on that PGU, but still
within the scheduling window. The third case is where a
failure is only observed after the current scheduling win-
dow has ended. By determining the energy produced in
each of these cases, the expected energy that each PGU
generates over the scheduling window may be calculated
and maximised.

The paper is organised as follows. A brief review of
the relevant literature is presented, comprising discussions
on reliability theory and GMS in §2. This is followed by
our novel GMS model formulation, exhibiting a nonlinear
objective function, in §3. In §4 the 21-unit test system
of Dahal and McDonald [3] and the 32-unit IEEE-RTS
[4] are reviewed briefly. In §5 we apply an exact solu-
tion methodology (implemented in the off-the-self mixed
integer programming solver IBM ILOG CPLEX) to an op-
timised piecewise linear approximation of the nonlinear
model of §3 within the context of the two aforementioned,
well-known GMS benchmark instances. In this manner,
the viability of our modelling approach is demonstrated.

The feasibility of our newly proposed GMS model is anal-
ysed in the form of a sensitivity analysis in §6 by relax-
ing the available manpower and scheduling window con-
straints. In the final section, §7 some concluding thoughts
are presented with respect to the anticipated feasibility
and effectiveness of our newly proposed GMS modelling
approach as an alternative to standard approaches in the
literature.

2. Literature review

This section contains a brief overview of certain basic
concepts in reliability theory, focusing on a description of
how to compute the reliability of a repairable system. The
section also contains a review of the literature on GMS
in which typical problem constraints are described along
with various popular optimality criteria and some typical
GMS model solution approaches applied in literature.

2.1. Reliability theory

Reliability theory, or survival theory as it is known in
statistics, is a general theory in engineering about system
failures. This theory consists of a number of ideas, models
and methods that may be employed to estimate, predict
and optimise the lifespan of a system [5]. One of the main
objectives in reliability theory is often to find a trade-off
between risking an unexpected failure and wasting a sys-
tem’s residual life. The two main branches of reliability
theory are theories that have been developed for repairable
systems and for non-repairable systems [6].

A system’s reliability $R(t)$ may be defined as the prob-
ability $P$ that its next failure time $T$ occurs at least $t$ time
units after its last failure. That is,

$$R(t) = P(T > t) = 1 - F(t) = \int_{t}^{\infty} f(\tau) d\tau, \quad (1)$$

where $f(t)$ is the probability density function (PDF) and
$F(t)$ the cumulative distribution function (CDF) of the
random variable $T$. The PDF $f(t)$ is also referred to as
the lifetime distribution model. This distribution is based
on the failure model selected to represent the system. In
the literature, a number of failure models are available, in-
cluding the non-homogeneous Poisson process (NHPP) fol-
lowing an exponential law [7], the NHPP following a power
law [8], and the homogeneous Poisson process (HPP) [9].

From these failure models, a model that best mimics the
failures of the system under consideration may be em-
ployed.
In this paper, the reliability of a PGU is calculated under the assumption that all the PGUs in the system under consideration are repairable systems. This assumption was also made by Wang and McDonald [10] due to the fact that a PGU typically spends most of its lifetime in the “useful life” phase of its so-called bath-tub curve followed by the failure rate function of the PGU components. Systems in this phase exhibit an approximately constant failure rate following an exponential distribution and therefore the failure model selected for modelling PGUs in this paper is the HPP failure model following an exponential law. The PDF of this failure model is

\[ f(t) = \lambda e^{-\lambda t}, \]  

(2)

where \( \lambda \) is failure rate of the lifetime distribution. The value of \( \lambda \) may be estimated from historical failure data of the system. Using (1), the probability that the next failure will be at least \( t \) time units from the present time may be calculated as

\[ R(t) = 1 - \int_0^t \lambda e^{-\lambda \tau} d\tau = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t}. \]  

(3)

The reliability \( R(t) \) in (3) represents the probability that a system will survive past time \( t \) and therefore is a value between 0 and 1, where \( t \) typically represents a global measure of system lifetime.

2.2. Generator maintenance scheduling

The aim in the GMS problem is to find a good planned maintenance schedule for all the PGUs in an entire power system [11] [12]. Various aspects should be taken into account when searching for good maintenance schedules. These aspects include satisfying the system energy demand and ensuring that the required number of maintenance personnel are available to perform maintenance during time periods when maintenance is scheduled. The adopted scheduling window in GMS models vary between a few weeks to several years, while the planning resolution is typically weekly [12], but daily, or even hourly, resolutions may also be found in the literature [11] [13] [14].

The GMS problem is seen as a complex, combinatorial optimisation problem which is generally tightly constrained. For this reason, instances of the GMS problem are hard to solve exactly for realistically sized input data sets [15], which has necessitated the use of approximate solution approaches such as heuristic approaches [14] [16] and metaheuristic approaches [13] [17] [18].

2.2.1. GMS model constraints

Model formulations for the GMS problem in the literature typically contain a variety of constraints in order to ensure adequate real-world representation. Perhaps one of the most important constraints is the energy demand constraint. This constraint ensures that during each period of the scheduling window, the forecasted system demand does not exceed the available system capacity of the PGUs [19]. It is therefore important to take cognisance of time periods of high energy demand as fewer PGUs should be scheduled for maintenance during these periods.

Maintenance window constraints specify earliest and latest times at which maintenance can be scheduled for each PGU in the system. These windows are typically difficult to establish and often are merely subjective rules implemented by the management of an energy system, or maintenance intervals specified by the suppliers of some of the larger components of PGUs [20].

In order to be able to perform maintenance on a PGU during a certain time period, certain resources are typically required. Resource availability constraints ensure that adequate levels of resources are available during all time periods when maintenance is scheduled for a PGU. This type of constraint typically focuses on the maintenance personnel required to perform the maintenance during a specific time period as in most models in the literature it is assumed that all the other resources will be available due to the maintenance being planned [12] [21].

Maintenance duration constraints specify the duration of maintenance, in units of the planning resolution, of con-
secutive time periods required to perform planned maintenance on each PGU. This type of constraint specifies the length of time during which each PGU in the system has to be offline when it is scheduled for planned maintenance [19]. To ensure that maintenance is performed over consecutive time periods, service contiguity constraints are also incorporated into most GMS models. These constraints specify that the planned maintenance of a PGU should be a contiguous period of time (i.e. should be uninterrupted) [19].

Some other examples of constraints that may be incorporated into GMS models are exclusion constraints and precedence constraints. Exclusion constraints specify certain sets of PGUs that are not allowed to be scheduled for maintenance simultaneously, whereas precedence constraints specify certain PGUs that should be scheduled for maintenance before certain other PGUs may be subjected to scheduling.

The aforementioned constraints can all either be modelled as soft or hard constraints. If a candidate solution violates a soft constraint, the solution is still considered feasible, but the objective function is penalised according to the number and degree of such constraint violations instead of being considered infeasible. Hard constraints, on the other hand, may not be violated by candidate solutions in order to be considered feasible [22].

2.2.2. GMS scheduling criteria

GMS model formulations may typically be categorised into three main classes of scheduling criteria: economic criteria, reliability criteria and convenience criteria [11, 23, 24]. Of these, convenience criteria are the least commonly employed for modelling purposes [11, 25, 23, 26]. Economic criteria come in two main incarnations, namely energy production cost minimisation and maintenance cost minimisation [11]. Due to the fact that maintenance cost is typically magnitudes smaller than energy production cost, the former cost often is ignored [14, 27]. Even though this is the mostly the case, there are some instances in the literature in which only the maintenance cost is minimised as scheduling criterion [28]. In some GMS models, the durations of PGU maintenance are allowed to vary within given limits. In these cases, maintenance cost is important, because it is linked with the appointment of additional maintenance personnel in terms of overtime work [11].

The class of reliability scheduling criteria is considered the most important class of GMS criteria among the three types of criteria [27, 24]. One of the most commonly adopted scheduling criteria in this class is levelling of the reserve generation margin over the entire operational scheduling window [11]. This is typically achieved by minimising the sum of squared reserves (SSR)

$$\left( \sum_{u \in U} C_{u,p} (1 - y_{u,p}) - D_p \right)^2,$$

where $y_{u,p}$ denotes the binary auxiliary variable which is 1 if planned maintenance of PGU $u$ in a set $U$ of PGUs in the power system is performed during planning period $p$ within a set $P$ of planning periods comprising the scheduling window, or zero otherwise. Furthermore, $C_{u,p}$ denotes the generating capacity of PGU $u \in U$ during time period $p \in P$ and $D_p$ denotes the expected demand of the system during planning period $p \in P$ [11, 29].

Minimising the loss of load probability (LOLP), minimising the expected unsupplied energy (EUE) and minimising the loss of load expectation (LOLE) are other popular criteria in the class of reliability scheduling criteria [30]. In order to calculate the LOLP, the expected probability of the system load exceeding its generation capacity is determined and minimised [24]. This probability is based on the distribution of the daily/weekly loads. Including the purchasing cost of additional energy due to unsupplied energy as a monetary value is a common approach towards quantifying the EUE scheduling criterion. With regards to the minimisation of LOLE, a similar approach is taken to that of the minimisation of LOLP, but the actual amount of time during which the system is not
able to meet the demand is taken as the objective function instead [31].

A more recent scheduling criterion introduced by Eygelaar et al. [32] to the class of reliability criteria is to minimise the probability of PGU failure over the duration of the scheduling window. This approach involves modelling constructs from the field of reliability theory to estimate the survivability of each of the PGUs in the system. This GMS criterion is also weighted according to the rated capacity of each PGU in order to give maintenance preference to PGUs that contribute more generating capability to the power system.

Finally, the class of convenience criteria, which is not often adopted in GMS models, mainly focuses on minimising deviations from a previously established PGU maintenance schedule [24]. These criteria usually entail finding ideal maintenance sequences of various PGUs, or minimising the violation of soft constraints in the model. Convenience criteria are seldom used in a single-objective modelling paradigm; they are typically employed in a multi-objective solution approach in conjunction with other, more important scheduling criteria [12, 24, 33].

2.2.3. Typical solution approaches

Four main approaches are typically employed in the literature to solve models of the GMS problem. A popular approach for small problem instances is the class of mathematical programming solution approaches which includes dynamic programming, as well as variations on the well-known branch-and-bound method [1]. As these are typically exact solution approaches, large computation times are usually required for their implementation [12].

Heuristic searches are methods aimed at uncovering not necessarily optimal, but at least good, solutions — often by applying a trial-and-error search approach. Although these methods do not always yield very good solutions, their strength is their relatively simple implementation and their very short computation times [16].

Expert systems embody extracted knowledge from experts in the field of GMS. This solution approach can often account for complicated factors or uncertainties, when scheduling planned maintenance, that other approaches are not always able to accommodate. When implementing expert systems, powerful and accurate rules are typically established in order to ensure that the systems are valid [1]. These rules then form part of the knowledge base of the system and are typically based on heuristic assumptions [14].

A popular solution approach for complicated medium to large GMS problem instances is the metaheuristic solution approach. Simulated annealing and genetic algorithms are the most popular algorithms in this class of solution approaches [31]. Other algorithms, such as tabu search [17], particle swarm optimisation [18, 35] and ant colony optimisation [19], have also been successfully employed in the literature to solve GMS problem instances.

Fuzzy systems is a fairly recent addition to the set of solution approaches employed to solve GMS problems. Fuzzy sets are employed to represent the objective functions and constraints of the GMS model, and are capable of accounting for some uncertainties in the model [36]. This class of solution approaches is typically employed in conjunction with other classes, such as dynamic programming techniques [14] and metaheuristic solution techniques designed to accommodate multiple objectives [37, 38], and have proved to return good solutions if acceptable computation times are allotted to them.

2.2.4. Recent developments

Until recently, the main focus of GMS research was on regulated systems, which are systems in which a single power utility is in charge of providing energy to an entire country. In regulated systems, on the other hand, the government of the country ensures that the power utility does not take advantage of the end user. Recent developments in GMS research have, however, shifted to deregulated sys-
tems where many power utilities exist and are able to produce and distribute energy within the same country, hence replacing monopolies with competition [15]. As deregulated systems thrive on competition, the majority of GMS problems involve cost-based or profit-based scheduling criteria as objective function in such cases.

Bisanovix et al. [39] presented a comprehensive model that forecasts energy prices on a weekly basis and takes into account long-term contracts with predefined power profiles. Game theory has also recently been employed, where the aim is to model coordination procedures for an independent system operator [40]. The Analytical Hierarchy Process has also been employed recently to incorporate the loss of power utility reputation and consumer loyalty towards a power utility. This was incorporated by Dahal et al. [41] as part of the opportunity cost of the solution. In 2017, Mazidi et al. [42] proposed a bi-level model which seeks to maximise the profit of the power utility, taking into account the reliability of the system as well as the cost limits of a society.

There have, however, also been some developments in the literature considering reliability-based scheduling objectives within the context of deregulated systems [15]. In a paper published in 2013, Elyas et al. [43] argued that power utilities are more competitive by being more reliable and therefore proposed two GMS objective functions for deregulated systems. The first objective seeks to maximise producers’ benefits and the reliability of the power grid, while the second objective maximises the annual social welfare. Some researchers have argued that the best GMS model for deregulated systems combines cost-based, profit-based and reliability-based GMS criteria in a multi-objective solution approach. Such a model was proposed by Zhan et al. [44], in which the scheduling criteria were maximising producer profit, maximising system reliability and minimising generation cost.

3. Mathematical problem formulation

In this section, we present our GMS model, including the novel objective function which accounts for the expected energy production for a system of PGUs, as well as the various model constraints applied.

3.1. The model variables

Suppose the power system contains $n$ PGUs indexed by the set $U = \{1, \ldots, n\}$ and that the GMS window is discretised into $m$ time periods of equal length, indexed by the set $P = \{1, \ldots, m\}$. Furthermore, let $x_{u,p}$ be a binary decision variable taking the value 1 if planned maintenance of PGU $u \in U$ is scheduled to start during time period $p \in P$, or zero otherwise. Let $y_{u,p}$ be a binary auxiliary variable taking the value 1 if planned maintenance of PGU $u \in U$ is scheduled to take place during time period $p \in P$, or zero otherwise.

A schematic representation of the variables defined above is provided in Figure 1, where $x'_u$ is a predetermined (negative) parameter denoting the time period during which PGU $u \in U$ entered into operation again after the previous maintenance operation performed on the PGU (during the previous scheduling window) and where $d_u$ is the duration of planned maintenance for PGU $u \in U$.

3.2. Objective function

Our novel objective function involves the maximisation of the expected system-wide energy production over the entire scheduling window, taking into account the possibility of unexpected failures of the PGUs in the system. We could not find any reference in the literature in which the expected energy production is maximised explicitly whilst taking into account unexpected failures of PGUs in the system. Such a maintenance criterion is expected to minimise the negative effects on power generation capabilities of the occurrence of unexpected failures of PGUs in the generating system. It is anticipated that a scheduling objective which maximises the system-wide expected energy
production may be very useful for power utilities as this is a measure of the total amount of available energy anticipated over a given scheduling window, which may aid in decision making when forecasting energy production levels for the given scheduling window.

Our GMS criterion takes into account three possible cases of failure occurrences of PGUs in the system under the assumption that each PGU will fail at most once during the maintenance scheduling window. The first case is where a PGU failure is observed before planned maintenance is performed on that PGU. The second case is where a PGU failure is observed after planned maintenance has already been performed on that PGU, but still within the scheduling window. The third case is where a failure is only observed after the current scheduling window has ended. The timing of PGU failures is modelled by the incorporation of random variables into our GMS model, facilitating calculation of the expected failure time for each PGU, employing methods from probability theory, reliability theory and the estimated failure rates for the PGUs.

In order to define the proposed scheduling criterion, three assumptions are made, namely that no PGU will fail more than once during the scheduling window, that each PGU will be subjected to maintenance exactly once during the scheduling window, and that the failures of different PGUs may be considered as independent events.

Let $X_u$ be a random variable denoting the timing of the next failure of PGU $u \in \mathcal{U}$ after commencement of the current scheduling window. Adopting the convention that the scheduling window is the interval $\tau = [0, T]$ on the real line, it follows that $X_u \in [0, \infty)$ for each PGU $u \in \mathcal{U}$. We assume that the PDF of the random variable $X_u$ takes the form

$$f(X_u) = \lambda_u e^{-\lambda_u X_u}, \quad (5)$$

where $\lambda_u$ is the failure rate of PGU $u \in \mathcal{U}$. The expected failure time $E(X_u)$ and the standard deviation $\sigma(X_u)$ of the inter-failure times for PGU $u \in \mathcal{U}$ are therefore both

$$E(X_u) = \sigma(X_u) = \frac{1}{\lambda_u}. \quad (6)$$

There are three cases to consider:

I A failure occurs before planned maintenance is performed on PGU $u \in \mathcal{U}$ and maintenance is either performed when the failure occurs or sometime thereafter, i.e. $0 \leq X_u \leq k_u$, where $k_u = \sum_{p \in \mathcal{P}} p x_{u,p} \in \mathcal{P}$ denote the time period during which maintenance is scheduled to commence within the scheduling window on PGU $u$.

II maintenance is performed before the next failure of PGU $u \in \mathcal{U}$, thus in effect postponing the failure to some time which is still within the scheduling window, i.e. $k_u < X_u \leq T$, or

III maintenance is performed before the next failure of PGU $u \in \mathcal{U}$, but the failure occurs after the scheduling window has ended, i.e. $T < X_u$.

The continuous variable $X_u$ may be interpreted within the discretised context of Figure 1. Let $n(k_u, X_u)$ denote the amount of energy generated by PGU $u \in \mathcal{U}$ during the scheduling window $\tau$ if PGU $u$ were to be subjected to

<table>
<thead>
<tr>
<th>Previous scheduling window</th>
<th>Current scheduling window</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{u,i-1}$</td>
<td>$x_{u,i}$</td>
</tr>
<tr>
<td>0</td>
<td>...</td>
</tr>
</tbody>
</table>
maintenance from period \(k_u \in \mathcal{P}\) to period \(k_u + d_u - 1\), given that a failure of PGU \(u \in \mathcal{U}\) occurs during period \(X_u\). Then the expected energy produced by PGU \(u \in \mathcal{U}\) during the scheduling window \(\tau\) may be determined as

\[
E(n(k_u, X_u)) = \int_0^{\infty} n(k_u, t)f(t)\,dt, \tag{7}
\]

where \(f(t)\) is the PDF of the failure model, specifying the probability \(P(X_u = t)\) of failure of PGU \(u \in \mathcal{U}\) at time \(t \in \tau\). The energy produced within the current scheduling window in the above three cases is

\[
n(k_u, X_u) = \begin{cases} \sum_{C \in \mathcal{U}} C_u k_u + C_u (T - (k_u + d_u - 1)), & \text{if } 0 \leq X_u \leq k_u \text{ (Case I)}, \\ C_u k_u + C_u (X_u - (k_u + d_u - 1)), & \text{if } k_u < X_u \leq T \text{ (Case II)}, \\ C_u k_u + C_u (T - (k_u + d_u - 1)), & \text{if } k_u < T < X_u \text{ (Case III)}, \end{cases} \tag{8}
\]

as illustrated in Figure 2. The expected energy produced by PGU \(u \in \mathcal{U}\) during the scheduling window is therefore obtained upon substitution of (8) into (7) as

\[
E(n(k, X_u)) = \sum_{u \in \mathcal{U}} \left[ C_u e^{-\lambda_u k_u} [(d_u - 1)\lambda_u - T\lambda_u - 1] + e^{\lambda_u k_u} (1 - \lambda_u ((d_u - 1) + k_u + T))] / \lambda_u \\
+ C_u \frac{1 + k_u \lambda_u + e^{-\lambda_u T -(k_u + (d_u - 1))]}{\lambda_u} \times ((d_u - 1)\lambda_u - T\lambda_u) - 1) \\
+ C_u e^{-\lambda_u T -(k_u + (d_u - 1))} [T - (d_u - 1)]. \tag{9}
\]

In Case I (the first term in (9)), energy is only available from the beginning of the scheduling window, when PGU \(u\) is assumed to be in operating condition, to the time when the failure occurs and then again from when scheduled maintenance has been completed until the end of the scheduling window, as shown in Figure 2(a).

In Case II (the second term in (9)), energy is only available from the beginning of the scheduling window, when the PGU is assumed to be in operating condition, to when maintenance is scheduled to start on PGU \(u\) and then again from the completion of the scheduled maintenance until when the failure occurs, as shown in Figure 2(b). As it is assumed that maintenance is only performed once during the current scheduling window, no energy will be generated by the PGU after the time at which the failure occurs (up to the end of the scheduling window) in Case II.

In Case III (the third term in (9)), energy is available from the beginning of the scheduling window, when the PGU is assumed to be in operating condition, to when maintenance is scheduled to start on PGU \(u\) and then from the completion of the scheduled maintenance until the end of the scheduling window, as shown in Figure 2(c).

The expected energy produced by the entire system of PGUs during the scheduling window is therefore

\[
\sum_{u \in \mathcal{U}} \left[ C_u e^{-\lambda_u k_u} [(d_u - 1)\lambda_u - T\lambda_u - 1] + e^{\lambda_u k_u} (1 - \lambda_u ((d_u - 1) + k_u + T))] / \lambda_u \\
+ C_u \frac{1 + k_u \lambda_u + e^{-\lambda_u T -(k_u + (d_u - 1))]}{\lambda_u} \times ((d_u - 1)\lambda_u - T\lambda_u) - 1) \\
+ C_u e^{-\lambda_u T -(k_u + (d_u - 1))} [T - (d_u - 1)]. \tag{10}
\]

which may be maximised as a maintenance scheduling criterion. This objective function is nonlinear, which makes it cumbersome to optimise exactly. Each term in the sum (10) may, however, be approximated by a piecewise linear function. This approximation process will result in the introduction of additional binary variables (and hence yield a larger model), but the benefit of the process is that it results in a linear optimisation model which is easier to solve exactly.

Two questions naturally arise naturally during this piecewise linearisation process: (1) What should the number of segments be in order to achieve a close approximation of a PGU expected energy production function? (2) Given that a suitable number of segments has been decided upon, what are the optimal positions for the internal breakpoints? The answer to the former question depends on the desired degree of closeness of the piecewise linear
approximation of the expected energy production function. Once the first question has, however, been answered, the second question is an optimisation problem that can be solved by means of dynamic programming.

The desirability of the locations of a given set of internal breakpoints may be quantified by assessing the piecewise linear approximation of the expected energy production function of a PGU by means of a regression model. The positioning of these breakpoints may be maximised by minimising the SSR of the aforementioned linear re-
gression model for each of the resulting line segments [45]. An algorithm for this purpose was developed by Bai and Perron [46, 45]. Solutions of this regression problem was previously considered by Bellman and Roth [47] as well as by Guthery [48]. Their work was, however, extended in 1997 by Bai and Perron [45] to accommodate multiple regression models and partial structural changes to the original model.

The method is initialised by constructing a $T \times T$ upper-triangular matrix of SSR for all the possible line segments in a piecewise linear approximation of a PGU’s expected energy production function. The rows of the matrix represent the possible starting dates corresponding to the line segment, while its columns represent the possible ending dates corresponding to the line segment. The entry in row $i$ and column $j$ of this matrix, denoted by $SSR(i, j)$, represents the SSR of an approximate line segment starting at time period $i$ and ending at time period $j$. The structure of such a triangular matrix is shown in Table 1.

The upper-triangular matrix in Table 1 may be constructed by means of a standard updating formula which calculates the recursive residuals on a segment-by-segment basis. Let $v(i, j)$ be the recursive residual at time $j$, using a sample of the observations starting at time $i$. Then the recursive relationship

$$SSR(i, j) = SSR(i, j - 1) + v(i, j)^2$$

holds [49], which may be used to populate the matrix. Once the matrix has been constructed with the relevant SSR contribution calculated for each line segment, a dynamic programming algorithm may be employed to evaluate which combination of line segments achieves a globally minimum value of the SSR.

Suppose the minimum length of an approximating line segment is $h$ and let $SSR\{P_{r,k}\}$ be the SSR associated with an optimal piecewise linear approximation containing $r$ internal breakpoints obtained by sampling the first $k$ observations in the data set. Then the global SSR for any number of line segments is a linear combination of the entries in the upper-triangular matrix of Table 1 [46]. An optimal partition is therefore a solution to the recursive problem

$$SSR\{\{P_{g,T}\} = \min_{g h \leq j \leq T - h} [SSR\{\{P_{g-1,j}\} + SSR(j, T)],$$

where $g$ is the number of line segments of the piecewise linear approximation. The procedure commences by evaluating all the sub-samples that allow one possible breakpoint ranging from observation $h$ to $T - gh$ in order to obtain a piecewise linear approximation with one internal breakpoint. During this step, the SSR of $T - (g - 1)h + 1$ optimal breakpoint partitions are calculated and stored. Each of these values has a corresponding ending date ranging from $2h$ and $T - (g - 1)h$ (inclusive). The next step is to obtain an optimal partitioning with two breakpoints, each of which each has a corresponding ending date ranging between $3h$ and $T - (g - 2)h$ (inclusive). This is achieved by considering all the possible one-breakpoint partitions in order to insert another breakpoint that will achieve a minimum SSR. In this way an optimal piecewise linear approximation containing two internal breakpoints is obtained. This forward recursive procedure is repeated until a set of $T - (g + 1)h + 1$ optimal internal breakpoints has been obtained, each of which has corresponding ending dates ranging from $(g - 1)h$ and $T - 2h$ (inclusive).

Although the algorithm executes very quickly, the majority of the computational time is attributed to the construction of the upper-triangular matrix of Table 1 [46]. The procedure may be implemented by invoking an R package called strucchange which contains a function called breakpoints. This function computes the optimal positions of the internal breakpoints, given a value of $g$, as described above [50].

### 3.3. Constraints

The GMS model proposed in this paper includes four classes of constraints. These constraint classes are energy
Table 1: The triangular matrix containing the SSR for line segments starting at date \(i\) and ending at date \(j\).

<table>
<thead>
<tr>
<th>Starting date</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>(\cdots)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>(\cdots)</th>
<th>(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SSR(1,1)</td>
<td>SSR(1,2)</td>
<td>SSR(1,3)</td>
<td>(\cdots)</td>
<td>SSR(1,T)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>SSR(2,2)</td>
<td>SSR(2,3)</td>
<td>(\cdots)</td>
<td></td>
<td>SSR(2,T)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>SSR(3,3)</td>
<td>(\cdots)</td>
<td></td>
<td></td>
<td>SSR(3,T)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\vdots)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SSR(T,T)</td>
<td></td>
</tr>
</tbody>
</table>

Demand satisfaction constraints (including specified safety margin), maintenance crew availability constraints, maintenance window constraints for each of the PGUs and exclusion constraints which specify that certain subsets of PGUs may not all be scheduled for simultaneous maintenance.

Let \(D_p\) denote the peak load demand required from the entire system of PGUs during time period \(p \in \mathcal{P}\), and let \(S\) denote the safety margin specified for the system (specified as a proportion of the total load demand of the system). The load demand constraint may then be formulated as

\[
D_p(1 + S) \leq \sum_{u=1}^{n} C_u(1 - y_{u,p}), \quad p \in \mathcal{P},
\]

Due to the fact that it is assumed that the resources required to perform planned maintenance are not necessarily the same during each period of maintenance, it is assumed that during the \(i\)-th time period of planned maintenance of a PGU \(u \in \mathcal{U}\), the PGU requires \(\Psi^u_i\) resources. Let \(\psi_{u,p,v}\) denote the resources required for planned maintenance on PGU \(u \in \mathcal{U}\) during time period \(p\) if planned maintenance of the unit starts during time period \(v \in \mathcal{P}\). The resource requirement parameters may then be calculated as

\[
\psi_{u,p,v} = \begin{cases} 
\Psi^u_{p-v+1}, & \text{if } 0 \leq p - v < d_u, \\
0, & \text{otherwise}.
\end{cases}
\]

Thereafter, the resource availability constraint may be formulated as

\[
\sum_{u=1}^{n} \sum_{v=1}^{p} \psi_{u,p,v} x_{u,v} \leq M_p, \quad p \in \mathcal{P},
\]

where \(M_p\) denotes the pre-specified number of available resources during time period \(p \in \mathcal{P}\) of the scheduling window.

Let \(e_u\) and \(\ell_u\) denote respectively the earliest starting time period and latest starting time for planned maintenance of each PGU \(u \in \mathcal{U}\). The maintenance window constraints are therefore formulated as

\[
\sum_{p=e_u}^{\ell_u} x_{u,p} = 1, \quad u \in \mathcal{U},
\]

under the assumption that planned maintenance of a PGU may only be scheduled once during the scheduling window. The maintenance window constraints may be formulated as

\[
x_{u,p} = 0, \quad \text{for all } p < e_u \text{ and all } p > \ell_u, \quad u \in \mathcal{U}.
\]

Let \(d_u\) denote the duration of planned maintenance on PGU \(u \in \mathcal{U}\). Then, clearly,

\[
y_{u,p} = 0, \quad \text{for all } p < e_u \text{ and all } p > \ell_u + d_u - 1, \quad u \in \mathcal{U}.
\]

The required duration of planned maintenance on each PGU is formulated as

\[
\sum_{p=e_u}^{\ell_u+d_u-1} y_{u,p} = d_u, \quad u \in \mathcal{U}.
\]

Linking constraints are included in our GMS model in order to enforce contiguity of the planned maintenance for the PGUs. These constraints are

\[
x_{u,p} \geq y_{u,p} - y_{u,p-1}, \quad u \in \mathcal{U}, \quad p \in \mathcal{P} \setminus \{1\},
\]

\[
x_{u,1} \geq y_{u,1}, \quad u \in \mathcal{U}
\]

(19)
and form a linkage between the auxiliary and decision
variables.

Exclusion constraints are also incorporated in the pro-
posed model which specify sets \( J_1, \ldots, J_w \) of PGUs that
are not allowed to be scheduled simultaneously for planned
maintenance. Let \( I_i \) denote the maximum number of PGUs
that are allowed to be scheduled for planned maintenance
simultaneously during any time period in exclusion set \( J_i \).
The exclusion constraints may then be formulated as
\[
\sum_{u \in J_i} y_{u,p} \leq I_i, \quad p \in \mathcal{P}, \quad i \in \{1, \ldots, w\}. \tag{20}
\]
The decision and auxiliary variables are of a binary nature;
therefore, the final constraints required are
\[
x_{u,p}, \ y_{u,p} \in \{0, 1\}, \quad u \in \mathcal{U}, \quad p \in \mathcal{P}. \tag{21}
\]

4. Test systems

The GMS model of §3 is solved within the context
of two well-known test systems from the GMS literature.
These two systems are the so-called 21-unit system, in-
troduced by Dahal and McDonald [3], and the IEEE-RTS
introduced by Schlinz and van Vuuren [4] in 2011, which
is based on the data set presented earlier by Albrecht
et al. [51]. This section is dedicated to a concise presenta-
tion of the data and model constraints specified for each
of these test systems.

4.1. The 21-unit test system

The 21-unit test test system contains twenty one PGUs
of which the specifications were presented in detail by Ya-
mayee [52]. In 1997, Dahal and McDonald [3] added some
additional constraints and certain simplifications to the
system presented by Yamaye [52]. The resulting system
exhibits a constant peak demand of 4,739 MW for each
week over the entire scheduling window which spans fifty
two weeks. Maintenance durations, as well as allowable
maintenance windows (i.e. earliest and latest maintenance
commencement dates), have also been specified for each
PGU in the power system. Furthermore, a maximum of
twenty maintenance personnel are available during each
week of the scheduling window to perform planned main-
tenance of the PGUs. The system updated by Dahal and
McDonald [3] does not, however, include the time peri-
ods during which the various PGUs were placed back in
operation upon completion of their last maintenance, nor
does it include the failure rates of the PGUs. The time
periods during which the various PGUs were placed back
in operation upon completion of their last maintenance
were therefore derived from a GMS solution proposed by
Schlinz and van Vuuren [4] for the 21-unit test system,
as presented in Table 2. The failure rates of the PGUs in
the system are presented in Table 3 — these values were
extrapolated from the RTS-79 system, presented in 1979
by Albrecht et al. [51], by comparing the sizes and main-
tenance durations of the PGUs in the two test systems.

<table>
<thead>
<tr>
<th>PGU</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x'_{u} )</td>
<td>−46</td>
<td>−15</td>
<td>−32</td>
<td>−50</td>
<td>−5</td>
<td>−49</td>
<td>−28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PGU</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x'_{u} )</td>
<td>−22</td>
<td>−39</td>
<td>−38</td>
<td>−51</td>
<td>−25</td>
<td>−40</td>
<td>−36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PGU</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x'_{u} )</td>
<td>−43</td>
<td>−30</td>
<td>0</td>
<td>−17</td>
<td>−10</td>
<td>−9</td>
<td>−46</td>
</tr>
</tbody>
</table>

In order to solve the nonlinear model described in §3
the objective function was linearised by means of the piece-
wise linear approximation method described in §3.2. The
breakpoints of the expected energy production curves pro-
duced by the various PGUs in the 21-unit test system
over the scheduling window are presented in Table 2. The
number of breakpoints in each approximation was selected
as the smallest value for which the corresponding piece-
wise linear function accounts for more than 99.5% of the
true expected energy production curve (in terms of the
Figure 3: The true curves representing the expected energy production of two PGUs in the 21-unit test system, together with the corresponding piecewise linear approximation of each curve. The curves represent the expected energy of the PGU with the largest number of breakpoints in Table 4 (PGU 17) as well as that of the PGU with the smallest number of breakpoints (PGU 19). The black dots indicate the corresponding breakpoints for each of the piecewise linear approximations.

Table 3: Assumed failure rates of the PGUs in the 21-unit test system.

<table>
<thead>
<tr>
<th>Unit</th>
<th>Failure rate</th>
<th>Unit</th>
<th>Failure rate</th>
<th>Unit</th>
<th>Failure rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0873</td>
<td>8</td>
<td>0.1178</td>
<td>15</td>
<td>0.1527</td>
</tr>
<tr>
<td>2</td>
<td>0.0873</td>
<td>9</td>
<td>0.1178</td>
<td>16</td>
<td>0.1527</td>
</tr>
<tr>
<td>3</td>
<td>0.0873</td>
<td>10</td>
<td>0.1178</td>
<td>17</td>
<td>0.0873</td>
</tr>
<tr>
<td>4</td>
<td>0.0873</td>
<td>11</td>
<td>0.1527</td>
<td>18</td>
<td>0.0873</td>
</tr>
<tr>
<td>5</td>
<td>0.0873</td>
<td>12</td>
<td>0.1527</td>
<td>19</td>
<td>0.3733</td>
</tr>
<tr>
<td>6</td>
<td>0.0873</td>
<td>13</td>
<td>0.0873</td>
<td>20</td>
<td>0.0873</td>
</tr>
<tr>
<td>7</td>
<td>0.0873</td>
<td>14</td>
<td>0.0873</td>
<td>21</td>
<td>0.0873</td>
</tr>
</tbody>
</table>

maximum deviation of the expected energy curve from the piecewise linear approximation function). The error proportion of the piecewise linear approximation function and the corresponding number of breakpoints are also presented in Table 4 for each PGU. The true expected energy production curve of PGU 17, for example, exhibiting the largest number of interior breakpoints, and that of PGU 19, exhibiting the smallest number of interior breakpoints, are shown in Figure 3. The piecewise linear approximation of each of these curves is also shown in the figure together with the corresponding number of breakpoints.

4.2. The 32-unit IEEE-RTS

The IEEE-RTS GMS benchmark instance was published in 1979 [51], and was initially referred to as the RTS-79 system. Since the first publication of this test system, revisions to the system were, however, published (in 1986 and again in 1999, referred to as the RTS-86 system [53] and the RTS-96 system [54], respectively).

A more recent revision of this system was published in 2011 by Schütz and van Vuuren [4], including a number of additional parameter values and constraints. This system is a larger, more complex system than the 21-unit test system and comprises load demand constraints (which include a safety margin), resource requirements and availability constraints, maintenance window constraints, and PGU maintenance exclusion constraints. The 2011 update of the GMS test instance is henceforth referred to as the IEEE-RTS and contains thirty two PGUs. The scheduling window again spans fifty two weeks within which each of the thirty two PGUs have to be scheduled exactly once.
Table 4: Optimal piecewise linear approximation breakpoints for the 21-unit test system which result in an expected energy production approximation remaining within an error band of 0.5% of the true expected energy production curve for each PGU.

<table>
<thead>
<tr>
<th>PGU</th>
<th># of breakpoints</th>
<th>Error %</th>
<th>Breakpoints</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>0.392</td>
<td>21, 35, 45</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.423</td>
<td>8, 19, 33, 44</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.346</td>
<td>15, 32, 44</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.409</td>
<td>27, 42</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.393</td>
<td>6, 13, 22, 33, 44</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0.484</td>
<td>27, 42</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>0.406</td>
<td>14, 31, 43</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>0.487</td>
<td>16, 35, 45</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>0.348</td>
<td>22, 34, 42, 48</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>0.275</td>
<td>31, 44</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>0.223</td>
<td>34, 45</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>0.425</td>
<td>33, 45</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>0.296</td>
<td>17, 33, 44</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>0.360</td>
<td>17, 33, 44</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>0.266</td>
<td>34, 45</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>0.315</td>
<td>33, 45</td>
</tr>
<tr>
<td>17</td>
<td>5</td>
<td>0.401</td>
<td>5, 11, 19, 30, 42</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>0.344</td>
<td>8, 19, 33, 44</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>0.209</td>
<td>46</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>0.487</td>
<td>7, 17, 30, 43</td>
</tr>
<tr>
<td>21</td>
<td>3</td>
<td>0.285</td>
<td>20, 35, 45</td>
</tr>
</tbody>
</table>

for planned maintenance. The capacities of the PGUs, the PGU maintenance durations, earliest and latest starting times for the maintenance of each PGU, and manpower requirements associated with the maintenance of each PGU (limited to twenty five during each week of the scheduling window) were all specified by Schlünnz and van Vuuren [4]. The IEEE-RTS [4] also contains a set of PGUs that form part of each of seven exclusion sets.

This system also requires a safety margin of 15% over and above the load demand which exhibits a typical two-seasonal peak demand characteristic. The peak load of the IEEE-RTS is reached during Week 51. As previously mentioned in [4][1], the time period during which each PGU was placed back in operation upon completion of its last maintenance procedure was not specified by Schlünnz and van Vuuren [4]. These instants were therefore again derived from the best solution proposed by Schlünnz and van Vuuren [4]. These time periods are presented in Table 5.

In the RTS-79 system [51, 53], the MTTF for each PGU was provided. These values were used to derive the failure rates of the PGUs in the system, as presented in Table 6.

Table 5: The times at which PGUs of the IEEE-RTS [4] are assumed to have been placed back in operation upon completion of their previous maintenance operations.

<table>
<thead>
<tr>
<th>PGU</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_u</td>
<td>−46</td>
<td>−27</td>
<td>−51</td>
<td>−8</td>
<td>−49</td>
<td>−21</td>
<td>−30</td>
</tr>
<tr>
<td>PGU</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>x_u</td>
<td>−16</td>
<td>−11</td>
<td>−25</td>
<td>−43</td>
<td>−38</td>
<td>−48</td>
<td>−15</td>
</tr>
<tr>
<td>PGU</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>x_u</td>
<td>−9</td>
<td>−26</td>
<td>−18</td>
<td>−18</td>
<td>−27</td>
<td>−37</td>
<td>−25</td>
</tr>
<tr>
<td>PGU</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>x_u</td>
<td>−44</td>
<td>−24</td>
<td>−18</td>
<td>−10</td>
<td>−26</td>
<td>−35</td>
<td>−29</td>
</tr>
<tr>
<td>PGU</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x_u</td>
<td>−16</td>
<td>−40</td>
<td>−33</td>
<td>−14</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Finally, the smallest number of breakpoint positions for which the piecewise linear approximation accounts for 99.5% of the true energy production curves are presented in Table 7 along with the error proportion of the approximation.

5. Experimental results

In this section, the results obtained when solving the piecewise linearised GMS model of §3 exactly within the context of the test systems of §4 are presented. An Intel Core™ i7-4770 processor with 8 GB RAM running at 3.4 GHz in a Microsoft™ Windows 7 64-bit operating system was used to perform all the computational results reported
Table 6: Assumed failure rates of the PGUs in the IEEE-RTS [51].

<table>
<thead>
<tr>
<th>Unit</th>
<th>Failure rate</th>
<th>Unit</th>
<th>Failure rate</th>
<th>Unit</th>
<th>Failure rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3733</td>
<td>12</td>
<td>0.1768</td>
<td>23</td>
<td>0.1527</td>
</tr>
<tr>
<td>2</td>
<td>0.3733</td>
<td>13</td>
<td>0.1768</td>
<td>24</td>
<td>0.0857</td>
</tr>
<tr>
<td>3</td>
<td>0.0873</td>
<td>14</td>
<td>0.1768</td>
<td>25</td>
<td>0.0857</td>
</tr>
<tr>
<td>4</td>
<td>0.0873</td>
<td>15</td>
<td>0.0571</td>
<td>26</td>
<td>0.0857</td>
</tr>
<tr>
<td>5</td>
<td>0.3733</td>
<td>16</td>
<td>0.0571</td>
<td>27</td>
<td>0.0857</td>
</tr>
<tr>
<td>6</td>
<td>0.0857</td>
<td>17</td>
<td>0.0571</td>
<td>28</td>
<td>0.0857</td>
</tr>
<tr>
<td>7</td>
<td>0.0857</td>
<td>18</td>
<td>0.0571</td>
<td>29</td>
<td>0.0857</td>
</tr>
<tr>
<td>8</td>
<td>0.1400</td>
<td>19</td>
<td>0.0571</td>
<td>30</td>
<td>0.1750</td>
</tr>
<tr>
<td>9</td>
<td>0.1400</td>
<td>20</td>
<td>0.1750</td>
<td>31</td>
<td>0.1750</td>
</tr>
<tr>
<td>10</td>
<td>0.1400</td>
<td>21</td>
<td>0.1750</td>
<td>32</td>
<td>0.1461</td>
</tr>
<tr>
<td>11</td>
<td>0.1400</td>
<td>22</td>
<td>0.1527</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Optimal piecewise linear approximation breakpoints for the IEEE-RTS which result in an expected energy production approximation remaining within an error band of 0.5% of the true expected energy production curve for each PGU.

<table>
<thead>
<tr>
<th>PGU</th>
<th># of breakpoints</th>
<th>Error %</th>
<th>Breakpoints</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.26</td>
<td>47</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.26</td>
<td>47</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.48</td>
<td>7, 16, 29, 42</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.48</td>
<td>8, 16, 29, 42</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.26</td>
<td>47</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.26</td>
<td>47</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>0.39</td>
<td>14, 31, 43</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>0.37</td>
<td>8, 19, 33, 44</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>0.40</td>
<td>6, 16, 34, 45</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>0.33</td>
<td>16, 36, 46</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>0.35</td>
<td>33, 45</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>0.34</td>
<td>36, 46</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>0.33</td>
<td>36, 46</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>0.44</td>
<td>10, 36, 46</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>0.40</td>
<td>8, 18, 29, 41</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>0.44</td>
<td>12, 27, 41</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td>0.33</td>
<td>9, 19, 31, 42</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>0.33</td>
<td>9, 19, 31, 42</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>0.43</td>
<td>12, 27, 41</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>0.35</td>
<td>36, 46</td>
</tr>
<tr>
<td>21</td>
<td>2</td>
<td>0.46</td>
<td>35, 46</td>
</tr>
<tr>
<td>22</td>
<td>2</td>
<td>0.50</td>
<td>34, 45</td>
</tr>
<tr>
<td>23</td>
<td>3</td>
<td>0.49</td>
<td>16, 37, 46</td>
</tr>
<tr>
<td>24</td>
<td>4</td>
<td>0.33</td>
<td>8, 19, 33, 44</td>
</tr>
<tr>
<td>25</td>
<td>4</td>
<td>0.43</td>
<td>7, 16, 29, 42</td>
</tr>
<tr>
<td>26</td>
<td>3</td>
<td>0.41</td>
<td>13, 30, 43</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>0.33</td>
<td>16, 33, 44</td>
</tr>
<tr>
<td>28</td>
<td>3</td>
<td>0.30</td>
<td>17, 33, 44</td>
</tr>
<tr>
<td>29</td>
<td>4</td>
<td>0.35</td>
<td>8, 19, 32, 43</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>0.34</td>
<td>36, 46</td>
</tr>
<tr>
<td>31</td>
<td>2</td>
<td>0.36</td>
<td>36, 46</td>
</tr>
<tr>
<td>32</td>
<td>4</td>
<td>0.40</td>
<td>8, 22, 38, 46</td>
</tr>
</tbody>
</table>
in the remainder of the paper. A well-known off-the-shelf software package, IBM ILOG CPLEX Optimization Studio [55, 56], was furthermore employed to solve the mathematical model presented in §3.

5.1. The 21-unit test system

An exact solution to the piecewise linearised GMS model of §3 was obtained for the 21-unit test system [3] by CPLEX within 36 seconds. The optimal decision variable values of this solution are given in integer decision vector form by $x = [11, 42, 13, 26, 33, 23, 20, 47, 16, 9, 6, 49, 4, 7, 2, 18, 30, 45, 52, 38, 15]$ which corresponds to an optimal objective function value of $219717$ MW·week (or 36912456 MWh). A graphical representation of the optimal maintenance schedule is presented as Scenario B in Figure 4(a). In Figure 5, the corresponding expected energy production for each of the PGUs is presented, with scheduled maintenance starting at the dates indicated by black dots.

The effects of taking into account the expected energy produced the system of PGUs over the scheduling window, as in the newly proposed objective function (10), may be analysed by comparing the results reported above to known solutions found in the literature. A popular...
(a) Comparison between the maintenance schedules of Scenarios A and B for the 21-unit test system

(b) The manpower required over the duration of the scheduling window for the two maintenance schedules in Figure 4(a)

(c) The available system capacity over the duration of the scheduling window for the two maintenance schedules in Figure 4(a)

Figure 4: Comparison between the maintenance schedules of Scenarios A and B for the 21-unit test system.
The optimal solution of Scenario B performs 18.66% worse than that of Scenario A in terms of the SSR scheduling objective, but it performs 10.29% better than that in Scenario A in terms of the newly proposed scheduling objective. The two objectives therefore conflict with each other in terms of when maintenance should be scheduled for the PGUs.

In order to minimise the SSR, maintenance schedules typically exhibit maintenance commencement dates that are spread out over the entire scheduling window. This is the case in Scenario A where maintenance on PGUs with large capacities is performed throughout the scheduling window. The scheduling objective in Scenario B, however, aims to schedule PGUs with large rated capacities
(PGUs that can produce more energy) close to the peaks of the expected energy curves of these PGUs (typically close to the end of the scheduling window). It may be seen in Figure 4(a) that PGUs with large rated capacities are typically scheduled for maintenance later, but within their respective PGU maintenance windows. PGU 19, for example, is scheduled for maintenance during the last week of the scheduling window in Scenario B because of exhibiting the highest failure rate of 0.3733, which is 10 weeks later than observed in Scenario A. A similar observation may be made for PGU 12, which is scheduled for maintenance 22 weeks later in Scenario B than in Scenario A due to its large failure rate of 0.1527.

The different effects of the two scheduling objectives may also be observed in Figure 4(b). In Scenario A, the manpower required for planned maintenance is spread out over the entire scheduling window. During each planning period of the problem instance, manpower is required as there is no planning period within the scheduling window during which no PGUs are in maintenance. In Scenario B, on the other hand, it is observed that towards the end of both halves of the scheduling window (e.g. weeks 1–27 and weeks 28–52) the manpower required is at the maximum available level as most of the PGUs exhibit peaks of their expected energy production curves towards the end of the scheduling window.

Similar observations may also be made in respect of Figure 4(c). In Scenario A, the available system capacity never falls below a 6% band above the demand (and safety margin). In Scenario B, on the other hand, the available system capacity drops down to 0.696% above the demand (and safety margin) during the middle stages of the scheduling window as PGUs with large capacities (e.g. PGUs 6, 7, 9) are scheduled for maintenance towards the end of the first half of the scheduling window (due to their particular maintenance window constraints). The maximum system capacity is mainly available towards the middle of the graph for both Scenarios A and B, whereas fewer discontinuities are observed in the available capacity for Scenario A than for Scenario B.

5.2. The 32-unit IEEE-RTS

A solution to the piecewise linearised GMS model of §3 was obtained for the IEEE-RTS [4] by CPLEX within 93 538 seconds (25.983 hours). The optimal decision variable values of this solution are given in integer decision vector form by $\mathbf{x} = [25, 12, 21, 30, 15, 51, 18, 33, 32, 48, 35, 14, 22, 42, 30, 38, 30, 31, 43, 23, 44, 17, 38, 36, 36, 49, 51, 51, 43, 13, 46, 38]$, which corresponds to an objective function value of 130 630 MW-week (or 21 945 840 MWh).

The effects of adopting a scheduling criterion that seeks to maximise the expected energy production, as in the newly proposed objective function of §3, may again be analysed by comparing the numerical results reported above with results found in the literature when adopting other reliability-related scheduling criteria, such as minimisation of the SSR. The numerical results reported above are therefore compared in this section with the results obtained by Schlünz and van Vuuren [4], who adopted the SSR scheduling criterion in (4). The results obtained by Schlünz and van Vuuren [4] (referred to here as Scenario C) are compared with the results reported above (referred to as Scenario D). The results of Schlünz and van Vuuren [4] for the IEEE-RTS (with the minimisation of SSR as objective) represent the best results available in the literature for this particular scheduling criterion and problem instance combination. Graphical representations of the two maintenance schedules are shown in Figure 6 (with the colour scale indicating the rated capacities of the PGUs in the system), while their corresponding effects on the manpower required and the available system capacity for the IEEE-RTS are shown in Figures 7(a) and 7(b) respectively.

It may be seen in Table 9 that the schedule of Scenario D performs 12.31% worse than that of Scenario C in terms of minimisation of the SSR objective. The schedule in
Scenario C, however, performs 19.67% worse in terms of the newly proposed scheduling objective than the schedule of Scenario D.

Table 9: Comparison between the objective function values associated with the maintenance schedules in Figure 6 for Scenarios C and D in the context of the IEEE-RTS. The percentage change values are computed for the solution of Scenario D relative to that of Scenario C.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>SSR (MW)$^2$</th>
<th>Expected energy production (MW-week)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>33 627 292</td>
<td>104 938</td>
</tr>
<tr>
<td>D</td>
<td>38 348 738</td>
<td>130 630</td>
</tr>
<tr>
<td>Percentage change</td>
<td>+12.31%</td>
<td>+19.67%</td>
</tr>
</tbody>
</table>

As previously mentioned, maintenance schedules typically exhibit maintenance commencement dates that are spread out over the entire scheduling window in order to minimise the SSR. This is indeed the case in Scenario C where maintenance is performed on PGUs with large capacities throughout the scheduling window. The objective in Scenario D, however, aims to schedule PGUs with large capacities close to the peak of the expected energy curves of these PGUs, which is typically close to the end of the scheduling window. It may be seen in Figure 6 that PGUs with large capacities are typically scheduled later, but still within their respective PGU maintenance windows. It may also be seen in Figure 6 that, compared to Scenario C, some PGUs with larger failure rates are either scheduled for planned maintenance early or late during the scheduling window in Scenario D. PGU 2, for example, is scheduled for maintenance during the early stages of the scheduling window due to exhibiting the largest failure rate of 0.3733, which is 13 weeks earlier than in Scenario C. On the other hand, it is observed for PGU 6, which also has a failure rate of 0.3733, that maintenance is scheduled 20 weeks later in Scenario D than in Scenario C.

The different effects of the two scheduling objectives on
manpower required and available system capacity may also be observed in Figure 7(a). In Scenario C, the manpower required for planned maintenance is spread out over the entire scheduling window. During each planning period of the problem instance, manpower is required as there is no planning period within the scheduling window during which no PGUs are in maintenance, until week 47. In Scenario D, on the other hand, it is observed that towards the end of both halves of the scheduling window (e.g., weeks 1–27 and weeks 28–52) the manpower required is at the maximum available level as most of the PGUs exhibit peaks in their expected energy production curves towards the end of the scheduling window.

Similar observations may also be made in respect of
6. Feasibility analysis

In this section, a feasibility analysis of the exact solution approach proposed in §5 for the model of §3 is presented within the context of the two test systems presented in §4 in the form of a sensitivity analysis involving various relaxations of the demand and maintenance scheduling window constraints.

6.1. The 21-unit test system

A piecewise linearisation approach towards solving the nonlinear GMS model of §3 for large power systems or very unconstrained systems (in terms of maintenance window constraints) is not expected to be feasible. The feasibility of an exact solution approach by CPLEX is influenced by the nature of the objective function (e.g. linear or nonlinear). It was demonstrated above that employing a piecewise linearisation model solution approach in the context of the 21-unit test system is feasible.

In order to analyse the effects of alterations in the system specifications on the feasibility of this piecewise linearisation model solution approach, six cases are analysed in this section in terms of the computation times required by CPLEX to solve the nonlinear model of §3. These cases involve combinations of increasing the peak demand of the system by a certain margin and relaxing the maintenance window constraints to have an earliest starting time of 1 and latest starting time of 53 less the duration of maintenance of each PGU. The first case is the original 21-unit test system which is considered as a reference case for the other five cases. The second case involves a 3% increase in the peak demand, but adheres to the original test system’s maintenance window constraints. The third case involves a 6.5% increase in the peak demand, but also adheres to the original maintenance window constraints. In the fourth case, the peak demand is kept as specified for the original 21-unit test system, but the maintenance window constraints are relaxed as described above. The fifth case involves a 3% increase in the peak demand and relaxed maintenance window constraints. Finally, the sixth case involves an increase in the peak demand of 6.5% and relaxed maintenance window constraints. Various statistics pertaining to optimal solutions to a piecewise linearisation of the nonlinear model in §3 are shown for these six cases in Table 10. Case 1 requires 36 seconds of computing time by CPLEX to obtain an optimal solution. In Case 2, the demand is increased by 3%, which requires 13 seconds of computing time and results in a 0.844% worsening of the objective function value. A decrease in computing time is observed and an optimal solution is still obtainable within a reasonable timeframe. Furthermore, when increasing the demand by 6.5%, in Case 3, a computation time of 27 seconds is required to obtain an optimal solution, which results in a 2.125% worsening of the objective function value compared to that in Case 1. It is observed that increasing the demand has a small impact on the computation time required to solve the nonlinear model (approximately) for the 21-unit test system. This computation time decreases by only 9 seconds over the course of a 6.5% increase in demand. The reason for the decrease in computation time as the demand is increased may be attributed to the smaller solution space through which the branch-and-cut algorithm has to search in order to obtain an optimal solution.
Table 10: Various statistics pertaining to optimal solutions to a piecewise linearisation of our GMS model in a sensitivity analysis in respect of demand and PGU maintenance windows for the 21-unit test system. The last column contains optimality gap values with respect to provable upper bounds on the maximum objective function value. An asterisk denotes that a time-out budget of 4 hours of computation time was reached by CPLEX.

<table>
<thead>
<tr>
<th>Case</th>
<th>Demand (%)</th>
<th>Maintenance window</th>
<th>Objective function value (MW-week)</th>
<th>Time (s)</th>
<th>Gap (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>Original</td>
<td>219 717</td>
<td>36</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>103</td>
<td>Original</td>
<td>217 862</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>106.5</td>
<td>Original</td>
<td>215 048</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>Relaxed</td>
<td>260 095*</td>
<td>14 400</td>
<td>0.83</td>
</tr>
<tr>
<td>5</td>
<td>103</td>
<td>Relaxed</td>
<td>255 114</td>
<td>5 963</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>106.5</td>
<td>Relaxed</td>
<td>243 566*</td>
<td>14 400</td>
<td>6.14</td>
</tr>
</tbody>
</table>

When the maintenance window constraints are relaxed, however, the number of feasible solutions to the nonlinear model increases dramatically, and so does the required computation time. Case 4, in which the demand is kept as specified for the original 21-unit test system and the maintenance window constraints are relaxed, results in a large increase in computation time. In fact, no optimal solution can be obtained within the time-out budget of 14 400 seconds of processing time. An optimality gap of 0.83% is achieved between the best objective function value (of 260 095) and the smallest provable upper bound on the objective function value within the allowed processing time. A decrease in computation time is observed in Case 5, where the demand is increased by 3% and the maintenance window constraints are relaxed. An optimal solution is obtained within 5 963 seconds of computation time and yields an objective function value that is 16.11% better than that of Case 1. Finally, in Case 6, it is again found that no optimal solution can be obtained within the time-out budget of 14 400 seconds of processing time. A gap of 6.14% is obtained between the best objective function value (of 243 566) and the smallest provable upper bound on the objective function value within four hours of computation time. It is therefore observed that relaxing the maintenance window constraints has a significant impact on the computation time required to solve the nonlinear model (approximately) for the 21-unit test system.

6.2. The 32-unit IEEE-RTS

It was demonstrated in §5.2 that employing a piecewise linearisation model solution approach in the context of the IEEE-RTS is not feasible within 8 hours of computation time as CPLEX required 25.983 hours to solve the model.

In order to analyse the effects of alterations in the system specifications on the piecewise linearisation model solution approach, four cases are analysed in this section in terms of the computation time required by CPLEX to solve the nonlinear model in the context of the IEEE-RTS. These cases again involve combinations of increasing the peak demand of the system by a certain margin and relaxing the maintenance window constraints to have an earliest starting time of 1 and latest starting time of 53 less the duration of maintenance of each PGU. The first case is the original IEEE-RTS, which is considered as a reference case for the other three cases. The second case involves a 3% increase in the peak demand, but adheres to the original test system’s maintenance window constraints. In the third case, the peak demand is kept as specified for the original IEEE-RTS, but the maintenance window constraints are relaxed as described above. Finally, the fourth
Table 11: Various statistics pertaining to optimal solutions to a piecewise linearisation of our GMS model in a sensitivity analysis in respect of demand and PGU maintenance windows for the IEEE-RTS. The last column contains gap values with respect to provable lower bounds on the maximum objective function value. An asterisk denotes that a time-out budget of 8 hours of computation time was reached by CPLEX.

<table>
<thead>
<tr>
<th>Case</th>
<th>Demand (%)</th>
<th>Maintenance window</th>
<th>Objective function value (MW-week)</th>
<th>Time (s)</th>
<th>Gap (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>Original</td>
<td>130 625</td>
<td>28 800</td>
<td>0.30</td>
</tr>
<tr>
<td>2</td>
<td>103</td>
<td>Original</td>
<td>128 590</td>
<td>28 800</td>
<td>0.45</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>Relaxed</td>
<td>142 169</td>
<td>28 800</td>
<td>3.16</td>
</tr>
<tr>
<td>4</td>
<td>103</td>
<td>Relaxed</td>
<td>138 633</td>
<td>28 800</td>
<td>3.20</td>
</tr>
</tbody>
</table>

In Case 1 it is found that no optimal solution can be obtained within the time-out budget of 28 800 seconds of computation time for the piecewise linearised model. The optimality gap achieved within the allowed budget of computation time for Case 1 was 0.3%. In Case 2, the demand is increased by 3%, which once again does not allow for the computation of an optimal solution as the maximum allowed computation time was reached. The best solution obtained yields a 1.558% worsening of the objective function value at a 0.45% optimality gap. Even with a smaller solution space to evaluate in Case 2, where the demand is increased by 3%, it is observed that the maximum computation time is again reached and the optimality gap also increases.

When the maintenance window constraints are relaxed, however, the number of optimal solutions to the piecewise linearisation of nonlinear model increases drastically, and so does the required computation time. Case 3, in which the demand is kept as specified for the original IEEE-RTS and the maintenance window constraints are relaxed, results in a large increase in the achievable optimality gap. Once again, no optimal solution can be obtained within the time-out budget of 28 800 seconds of computation time. An optimality gap of 3.16% is obtained between the best objective function value and the smallest provable upper bound on the objective function value within the allowed processing time. This results in a 90.51% increase in the optimality gap relative to that of Case 1. Finally, as was observed for the previous three cases, no optimal solution is achievable within the time-out budget of 28 800 seconds of computation time. In Case 4, the demand is increased by 3% and the maintenance window constraints are also relaxed. The best solution obtained within the allowed computation time exhibits an objective function value of 138 633 with an optimality gap of 3.20% between this objective function value and the smallest provable upper bound on the objective function value within the allowed processing time.

7. Conclusion

In this paper, a new scheduling criterion was proposed for PGU maintenance scheduling. The proposed criterion involves maximising the expected energy produced by all the PGUs in the power system. Our GMS objective function incorporates a random variable which represents the possibility of a failure being observed for each PGU. Three cases were considered for this random variable. The first is where a failure is observed before maintenance is performed on a PGU, the second is where a failure is observed after maintenance has been performed, but before the end
of the scheduling window, and the final case is where a failure is observed after the scheduling window has ended.

A GMS model was derived based on the criterion mentioned above. The proposed objective function is, however, nonlinear and the model therefore cannot be solved exactly by an off-the-shelf solvers. The scheduling criterion was consequently linearised by employing a piecewise linear approximation approach, selecting the smallest number of breakpoints such that 99.5% of the true energy curve is represented. The optimal positioning of these breakpoints was obtained by employing dynamic programming.

This piecewise linear solution approach was employed to solve two well-known benchmark instances in the GMS literature. The first system is the 21-unit system which was introduced by Dahal and McDonald [3] and contains twenty one PGUs that each has to be scheduled for maintenance exactly once within a scheduling window of fifty two weeks. The second benchmark system was introduced by Schlünz and van Vuuren [4] and is based on the IEEE-RTS [4] data set [51, 53]. This system contains thirty two PGUs that also each has to be scheduled exactly once for maintenance within a scheduling window of fifty two weeks.

The optimal solutions obtained for the piecewise linearisation models were compared to GMS schedules proposed for these two benchmark systems in the literature aimed at minimisation of the SSR. A sensitivity analysis was finally performed in order to analyse the effects of various combinations of relaxations of the demand and maintenance scheduling window constraints for both the aforementioned test systems. A maximum of 14 400 seconds of computation time was allowed for the smaller 21-unit test system whereas 28 800 seconds was allowed for the larger IEEE-RTS. An optimal solution was found within the allowable computation time for the 21-unit test system whereas 28 800 seconds was allowed for the larger IEEE-RTS. An optimal solution was found within the allowable computation time for the 21-unit test system, but in two cases of the subsequent feasibility sensitivity analysis the time-out was reached without an optimal solution being obtained. With regards to the IEEE-RTS, it was observed that an optimal solution could not be obtained within the allowable computation time for any of the cases considered in the sensitivity analysis. It is therefore anticipated that adopting an exact solution approach toward solving a piecewise linearisation of the GMS model of §3 for a real-world problem instance, which may easily contain more than a hundred PGUs, will not be practically feasible within a short computation time frame. We therefore propose that in such cases the model of §3 should be solved approximately by means of an appropriate meta-heuristic.

References


[56] IBM Corporation . IBM ILOG CPLEX Optimizer Studio. 2016. URL: https://goo.gl/t0AMdb