A Bi-Objective Model for Nurse Shift Scheduling under a Human-Factors (Fatigue) Approach

Mathematical Model and Solution Methodology

1 Overview

This document presents a mathematical optimization model for scheduling nursing staff across the shifts of a hospital ward, together with the metaheuristic procedure developed to solve it. The distinguishing feature of the model is that it embeds *human-factors engineering* directly into the optimization: rather than minimizing cost alone, it treats **nurse fatigue** as a second, competing objective and models the recovery effect of scheduled rest breaks using the *virtual-age* concept borrowed from reliability and maintenance theory.

The result is a **bi-objective stochastic scheduling problem**. Because the two objectives—total cost and total fatigue—conflict, there is no single optimum; instead the goal is to recover the set of non-dominated (Pareto-optimal) trade-off solutions. The problem is solved by a hierarchical metaheuristic combining a genetic-algorithm encoding with a variable-neighborhood search (referred to throughout as GA–VNS), and the quality of its output is validated against exhaustive enumeration on small instances.

2 Problem Definition

The planning horizon is a 24-hour day partitioned into three consecutive shifts (morning, evening, night), indexed by $i = 1, \ldots, T$ with $T = 3$. For each shift the planner decides how many nurses to add or remove relative to the previous shift, and how to schedule rest breaks (their spacing and recovery depth) so as to control accumulated fatigue.

Demand for nurses in each shift is uncertain: the number of patients (and hence required nurses) is a random variable that follows a known statistical distribution. Staffing too few nurses incurs a *shortage* penalty; staffing too many incurs a *surplus* penalty. Hiring and releasing nurses between shifts also carries a cost.

2.1 Modeling Assumptions

1. Only a single nurse type (homogeneous staff) is considered.
2. Fatigue is a strictly increasing function of (virtual) time worked.
3. Rest breaks may be taken only at discrete points in time.
4. Each rest break reduces fatigue, but *not* back to the start-of-shift level (imperfect recovery).
5. At the start of each shift a nurse begins at the minimum fatigue level.
6. Demand for nurses is random with a known probability distribution (taken as Poisson).
7. Staffing is revised every shift; a nurse serves in exactly one shift.
8. Rest duration is negligible relative to the shift length.

3 Notation

Symbol	**Definition**
Indices and sets	
i	Shift (period) index, $i = 1, \dots, T$
j	Rest-break index within a shift, $j = 1, \dots, n_i$
Cost parameters	
c_1	Cost of hiring (adding) one nurse
c_2	Cost of releasing (removing) one nurse
c_3	Unit shortage cost (per missing nurse)
c_4	Unit surplus cost (per excess nurse)
C_0	Wage per nurse per hour
a, γ	Positive shape parameters of the fatigue-reduction cost
Demand and fatigue	
D	Random demand (required nurses) in a shift
λ_i	Poisson arrival rate of patients in shift i
L_i	Duration of shift i
$F(\cdot)$	Fatigue function (of virtual age)
b	Maximum admissible fatigue level
β, θ	Weibull shape and scale parameters of the fatigue function
Decision and state variables	
x_i	Number of nurses on duty in shift i
x_i^+	Number of nurses added entering shift i
x_i^-	Number of nurses removed entering shift i
m_i	Rest (recovery) level applied in shift i
Δ_i	Interval between consecutive rest breaks in shift i
n_i	Number of rest breaks in shift i, $n_i = L_i/\Delta_i$
τ_{ij}	Time at which the j-th rest break is applied in shift i
$\delta(m_i)$	Rejuvenation function, $\delta(m_i) \in [0, 1]$
$v_i(t)$	Virtual age of a nurse at time t in shift i
CPM_i	Fatigue-reduction (preventive-maintenance) cost in shift i

4 Mathematical Model

4.1 Staffing Balance and Adjustment Cost

Let x_{i-1} be the number of nurses in the previous shift. Entering shift i the level may be raised by x_i^+ or lowered by x_i^-, giving the balance

$$x_i = x_{i-1} + x_i^+ - x_i^-, \qquad i = 1, \dots, T, \tag{1}$$

where additions and removals are mutually exclusive in any single shift:

$$x_i^+ \cdot x_i^- = 0, \qquad i = 1, \ldots, T. \tag{2}$$

With c_1 and c_2 the unit hiring and release costs, the total adjustment cost over the horizon is

$$c_1 \sum_{i=1}^{T} x_i^+ \; + \; c_2 \sum_{i=1}^{T} x_i^-. \tag{3}$$

4.2 Stochastic Demand: Expected Shortage and Surplus

Demand D in shift i is a Poisson random variable with rate λ_i, so that

$$\Pr(D = d) = \frac{e^{-\lambda_i}\,\lambda_i^d}{d!}, \qquad d = 0, 1, 2, \ldots \tag{4}$$

A *shortage* occurs when demand exceeds the staffed level ($D > x_i$) and a *surplus* occurs when staff exceeds demand ($D \leq x_i$). The expected shortage and expected surplus in shift i are, respectively,

$$\mathbb{E}[\text{shortage}_i] = \sum_{D=x_i}^{D_{\max}} (D - x_i)\, \frac{e^{-\lambda_i}\lambda_i^D}{D!}, \tag{5}$$

$$\mathbb{E}[\text{surplus}_i] = \sum_{D=0}^{x_i} (x_i - D)\, \frac{e^{-\lambda_i}\lambda_i^D}{D!}. \tag{6}$$

Weighting by the unit shortage cost c_3 and unit surplus cost c_4 and summing over all shifts gives the expected mismatch cost,

$$c_3 \sum_{i=1}^{T} \sum_{D=x_i}^{D_{\max}} (D - x_i)\, \frac{e^{-\lambda_i}\lambda_i^D}{D!} \; + \; c_4 \sum_{i=1}^{T} \sum_{D=0}^{x_i} (x_i - D)\, \frac{e^{-\lambda_i}\lambda_i^D}{D!}. \tag{7}$$

4.3 Fatigue via Virtual Age (Imperfect Maintenance)

The model treats a rest break like a *preventive-maintenance* action on a nurse. In reliability theory a maintenance action leaves a component somewhere between two bounds: *as good as new* (fully restored) and *as bad as old* (no improvement); the realistic intermediate case is *imperfect maintenance*. The degree of recovery is captured by the rejuvenation function

$$\delta(m_i) \in [0, 1], \qquad \begin{cases} \delta(m_i) = 0 & \text{(as good as new)} \\ \delta(m_i) = 1 & \text{(as bad as old)} \\ 0 < \delta(m_i) < 1 & \text{(imperfect recovery).} \end{cases} \tag{8}$$

Rest breaks are applied at times $\tau_{i1}, \tau_{i2}, \ldots, \tau_{in_i}$ spaced by a fixed interval Δ, so the number of breaks in a shift of length L_i is $n_i = L_i/\Delta$. The *virtual age* accumulates real elapsed time during work and is partially reset at each break. After the j-th break,

$$v_{ij} = v_{i,j-1} + \delta(m_i)\left(\tau_{ij} - \tau_{i,j-1}\right), \qquad v_{i0} = 0, \tag{9}$$

and at an arbitrary instant t between two breaks,

$$v_i(t) = v_{i,j-1} + \left(t - \tau_{i,j-1}\right), \qquad \tau_{i,j-1} \le t < \tau_{ij}, \quad j = 2, \ldots, n_i. \tag{10}$$

The instantaneous fatigue is obtained by evaluating the fatigue function F at the virtual age, subject to the cap b:

$$F\big(v_i(t)\big) = \begin{cases} F(t) & t < \tau_{i1},\ F(t) < b, \\ F\big(v_{i,j-1} + t - \tau_{i,j-1}\big) & \tau_{i,j-1} \le t < \tau_{ij},\ F < b, \\ b & F \ge b. \end{cases} \tag{11}$$

The fatigue function itself is taken in Weibull form (shape β, scale θ), consistent with a hazard that increases with virtual age. The *average fatigue* a single nurse accumulates over shift i is the area under the fatigue curve,

$$\mathbb{E}[F_i] = \int_0^{\tau_{i1}} F(t)\, dt \;+\; \sum_{j=2}^{n_i} \int_{\tau_{i,j-1}}^{\tau_{ij}} F\big(v_{ij} + t - \tau_{i,j-1}\big)\, dt. \tag{12}$$

4.4 Fatigue-Reduction Cost

Scheduling rest breaks to mitigate fatigue is not free; the associated cost in shift i is

$$\mathrm{CPM}_i = m_i\, \Delta_i\, C_0\, a^{\,m_i}\, \gamma, \tag{13}$$

where C_0 is the hourly wage, m_i the rest level, Δ_i the inter-break interval, and $a, \gamma > 0$ are tuning parameters. This term couples the fatigue-management decisions back into the cost objective.

4.5 Complete Bi-Objective Model

Collecting the components, the model minimizes total cost (f_1) and total fatigue (f_2) simultaneously:

$$\begin{aligned} \min\ f_1 = {} & \sum_{i=1}^{T} \mathrm{CPM}_i + c_1 \sum_{i=1}^{T} x_i^+ + c_2 \sum_{i=1}^{T} x_i^- \\ & + c_3 \sum_{i=1}^{T} \sum_{D=x_i}^{D_{\max}} (D - x_i) \frac{e^{-\lambda_i} \lambda_i^D}{D!} + c_4 \sum_{i=1}^{T} \sum_{D=0}^{x_i} (x_i - D) \frac{e^{-\lambda_i} \lambda_i^D}{D!} \end{aligned} \tag{14}$$

$$\min\ f_2 = \sum_{i=1}^{T} x_i \left[\int_0^{\tau_{i1}} F(t)\, dt + \sum_{j=2}^{n_i} \int_{\tau_{i,j-1}}^{\tau_{ij}} F\big(v_{ij} + t - \tau_{i,j-1}\big) dt + \int_{\tau_{in_i}}^{L_i} F\big(v_{in_i} + t - \tau_{in_i}\big) dt \right] \tag{15}$$

subject to

$$v_{ij} = v_{i,j-1} + \delta(m_i)\big(\tau_{ij} - \tau_{i,j-1}\big), \qquad j = 2, \dots, n_i, \tag{16}$$
$$v_i(t) = v_{i,j-1} + \big(t - \tau_{i,j-1}\big), \qquad \tau_{i,j-1} \le t < \tau_{ij}, \tag{17}$$
$$x_i = x_{i-1} + x_i^+ - x_i^-, \qquad i = 1, \dots, T, \tag{18}$$
$$x_i^+ \cdot x_i^- = 0, \qquad i = 1, \dots, T, \tag{19}$$
$$x_{\min} \le x_i \le x_{\max}, \qquad i = 1, \dots, T, \tag{20}$$
$$x_i^+, x_i^-, m_i, \Delta_i \ge 0, \text{ integer where applicable.} \tag{21}$$

The first objective minimizes the sum of staffing-adjustment costs, expected shortage and surplus costs, and the cost of the fatigue-mitigation policy. The second objective minimizes the total expected fatigue accumulated by all nurses across all shifts. The constraints enforce the virtual-age dynamics, the shift-to-shift staffing balance, and the bounds on the staffing level.

5 Solution Methodology

5.1 Rationale

Because f_1 and f_2 pull in opposite directions—leaner staffing and fewer breaks lower cost but raise fatigue, while more staff and more rest lower fatigue but raise cost—no single solution optimizes both. The aim is therefore to construct the **Pareto front**: the set of non-dominated solutions for which no objective can be improved without worsening the other. Two solvers are used. A hierarchical GA–VNS metaheuristic generates near-optimal Pareto sets efficiently for medium and large instances, and exhaustive enumeration provides an exact reference on small instances for validation.

5.2 Solution Encoding

A candidate solution is represented as a 4×3 matrix: one column per shift and one row per decision family $\{x_i^+, x_i^-, m_i, \Delta_i\}$. For example,

	shift 1	shift 2	shift 3
x_i^+	2	0	0
x_i^-	0	1	1
m_i	1	2	1
Δ_i	·	·	·

In the reported results each Pareto solution is stored as a flattened 1×15 vector whose five triples give, in order, the nurses hired, the nurses released, the rest interval, the fatigue level, and the resulting staffing state per shift.

5.3 The Hierarchical GA–VNS Algorithm

The procedure builds the dominated-solution (Pareto) set through six steps.

Step 1 — Representation. Map the problem from its natural space to the coded 4×3 matrix described above.

Step 2 — Initial solution. Generate a random feasible solution. For each shift, one of the addition/removal rows is chosen at random and filled with a random integer in $[x_{\min}, x_{\max}]$; if that value is positive the complementary row is set to zero (and vice versa). Rows for m_i and Δ_i are filled with random integers in $[m_{\min}, m_{\max}]$ and $[\Delta_{\min}, \Delta_{\max}]$.

Step 3 — Fitness. Evaluate f_1 and f_2 for the candidate and insert it into the Pareto set.

Step 4 — Neighborhood moves. Generate new candidates using six neighborhood structures, each perturbing one cell by one unit within its bounds:

N1. Pick a random column of the add/remove rows; *increment* the staffing change if $x_{ij} + 1 \leq x_{\max}$, then zero the complementary cell.

N2. As N1 but *decrement* the staffing change if $x_{ij} - 1 \geq x_{\min}$.

N3. Pick a random column of the rest-level row; *increment* m_j if $m_j + 1 \leq m_{\max}$.

N4. As N3 but *decrement* m_j if $m_j - 1 \geq m_{\min}$.

N5. Pick a random column of the interval row; *increment* Δ_j if $\Delta_j + 1 \leq \Delta_{\max}$.

N6. As N5 but *decrement* Δ_j if $\Delta_j - 1 \geq \Delta_{\min}$.

The search starts from N1. If a move yields a solution that enters the Pareto set the search stays in the current structure; otherwise it advances to the next structure. Whenever a new solution is accepted into the Pareto set, the search resets to N1. This stage performs M search iterations.

Step 5 — Pareto update. A new solution is evaluated against the current Pareto set under three cases:

- **Case 1.** If the new solution is worse in both objectives (or worse in one and equal in the other) than every member, it is rejected.
- **Case 2.** If it strictly dominates k members in both f_1 and f_2, it is admitted and those k dominated members are removed.
- **Case 3.** If it does not strictly dominate any member but is strictly better than all members in one objective, it is admitted and no member is removed.

Step 6 — Termination. Repeat Steps 2–4 up to a maximum iteration count.

5.4 Exhaustive-Search Baseline

For small instances the entire feasible space is enumerated to obtain the exact Pareto set. This serves as ground truth: the GA–VNS Pareto set is compared against it through a similarity rate, denoted **SRate**, defined as the percentage overlap between the two fronts.

6 Computational Experiments and Results

6.1 Experimental Setup

The model was implemented in MATLAB R2016 and executed on a desktop machine (Intel G3250 at 3.20 GHz, 4 GB RAM). The key parameter settings were as follows.

Parameter	Value
Poisson arrival rates $(\lambda_1, \lambda_2, \lambda_3)$	(2, 3, 3)
Weibull fatigue parameters (β, θ)	(2, 500)
Mutation rate	0.1
Crossover rate	0.9
Iterations	200
Hourly wage levels C_0	8000, 10000, 12000, 14000, 16000
Starting staffing states	9 configurations, e.g. $[6,3,3], [5,3,3], [4,3,3], [5,2,3], \ldots$

Each of the nine starting configurations was solved at each of the five wage levels by both methods, recording the Pareto set, the Pareto fitness values, the runtime, and the SRate.

6.2 Findings

1. The metaheuristic produces near-optimal fronts. Across every reported instance the similarity between the GA–VNS Pareto set and the exact exhaustive-search set (**SRate**) remained above 90%, frequently in the 94–98% range. This supports the conclusion that the GA–VNS solutions are acceptable approximations of the true Pareto front.

2. The metaheuristic is dramatically faster. GA–VNS solved each instance in approximately 1–3 seconds, whereas exhaustive search required from roughly 30 seconds to over 500 seconds depending on instance size. The metaheuristic thus delivers $\sim$90–98% of the solution quality at a small fraction of the computational effort—the trade-off that makes it the practical choice for larger problems where enumeration is infeasible.

3. Lower wages shift the entire frontier favorably. As the hourly wage C_0 decreases, the resulting Pareto front *dominates* the fronts obtained at higher wages, because lower wages permit greater savings in both staffing-provision and fatigue-reduction costs, moving the whole cost–fatigue trade-off curve to a better position.

7 Summary

The problem was formulated as a bi-objective stochastic nurse-rostering model that, uniquely, incorporates human fatigue as a first-class objective. Fatigue is modeled through the virtual-age / imperfect-maintenance framework, in which scheduled rest breaks act as partial recovery actions whose depth is governed by a rejuvenation function. The first objective aggregates staffing-adjustment, expected shortage/surplus (under Poisson demand), and fatigue-mitigation costs; the second aggregates total expected fatigue (with a Weibull fatigue function) across all nurses and shifts. Because the two objectives conflict, the solution is a Pareto front rather than a single point. A hierarchical GA–VNS metaheuristic recovers this front efficiently, and validation against exhaustive enumeration shows it attains better than 90% agreement with the exact front while running one-to-three orders of magnitude faster.