

Hidden Markov Chains - Project report

Noé AMAR, Wissal HAOUAMI, Beatriz EVELBAUER, Alexandre HEYMANN

January 2026

Contents

1	Model	2
1.1	Data and notation	2
1.2	Latent team strengths (state process)	2
1.3	Observation model for match outcomes	2
1.4	Parameter vector and identifiability	3
1.5	Inference targets: filtering and rankings	3
2	Sequential Monte Carlo Inference	4
2.1	Filtering objective	4
2.2	Decoupling (factorial) approximation	4
2.3	Particle representation	4
2.4	Bootstrap particle filter	5
2.5	Log-likelihood estimation	5
2.6	Discussion	6
3	Parameter Estimation via EM	6
3.1	Identifiability and parameter selection	6
3.2	Monte Carlo EM principle	6
3.3	Log-likelihood surface	7
3.4	EM trajectories	7
4	Results	7
4.1	Filtering results and team strength trajectories	8
4.2	Log-likelihood surface and parameter estimation	8
5	Discussion and Limitations	10
6	Conclusion	11

1 Model

1.1 Data and notation

We consider football matches from the English Premier League over the seasons 2018–2019 to 2022–2023. Each match is played between two teams: a home team and an away team. We order all matches by their match time and index them by $k = 1, \dots, K$ with non-decreasing times $0 = t_0 \leq t_1 \leq \dots \leq t_K$. Let N be the number of distinct teams appearing in the dataset across the considered seasons.

For match k , we denote by $h(k) \in [N]$ the index of the home team, by $a(k) \in [N]$ the index of the away team, and by $y_k \in \mathcal{Y}$ the observed outcome, with

$$\mathcal{Y} = \{\text{H}, \text{D}, \text{A}\},$$

corresponding respectively to home win, draw, and away win.

1.2 Latent team strengths (state process)

We model each team $i \in [N]$ with a latent, time-varying strength process $\{x_t^i\}_{t \geq 0}$ taking values in R . Higher values correspond to stronger teams.

Following the state-space modelling perspective for skill rating, we assume: (i) team strengths are independent a priori across teams, (ii) each strength evolves as a Markov process in continuous time, (iii) observations (match outcomes) depend only on the two teams involved in the match. These assumptions match the general formulation presented in the reference paper. [?]

Initial distribution. At time $t = 0$, each team strength is initialized independently as

$$x_0^i \sim \mathcal{N}(0, \sigma_0^2), \quad i = 1, \dots, N,$$

where $\sigma_0 > 0$ controls the prior uncertainty about team strength at the start of the observation window.

Dynamics. Between match times, team strengths evolve according to a continuous-time Gaussian random walk (Brownian motion), so that for any $t < t'$,

$$x_{t'}^i \mid x_t^i \sim \mathcal{N}(x_t^i, \tau^2(t' - t)),$$

where $\tau > 0$ governs how quickly strengths drift over time. In particular, when τ is small, team strengths are close to static; when τ is large, strengths can change rapidly.

1.3 Observation model for match outcomes

For match k , define the latent skill difference

$$d_k = x_{t_k}^{h(k)} - x_{t_k}^{a(k)}.$$

We adopt a three-outcome (home/draw/away) likelihood based on a logistic sigmoid, as commonly used in paired-comparison models and as discussed in the paper’s state-space framework for sports with draws. [?]

Let $\sigma(\cdot)$ denote the logistic sigmoid $\sigma(z) = (1 + e^{-z})^{-1}$. We introduce:

- a scale parameter $s > 0$ controlling how sensitive outcomes are to skill differences,
- a draw parameter $\varepsilon \geq 0$ controlling how frequent draws are.

Then the outcome probabilities are defined as

$$P(y_k = \text{D} \mid d_k) = \sigma\left(\frac{d_k + \varepsilon}{s}\right) - \sigma\left(\frac{d_k - \varepsilon}{s}\right),$$

$$P(y_k = \text{H} \mid d_k) = \sigma\left(\frac{d_k - \varepsilon}{s}\right), \quad P(y_k = \text{A} \mid d_k) = 1 - \sigma\left(\frac{d_k + \varepsilon}{s}\right).$$

This parameterization ensures that probabilities are non-negative and sum to 1.

1.4 Parameter vector and identifiability

Collect the static parameters as

$$\theta = (\sigma_0, \tau, s, \varepsilon).$$

The model is invariant to additive shifts of all strengths (only differences matter), which we fix by centering the prior mean at 0. There is also a scale non-identifiability between (s, τ, σ_0) in the sense that rescaling latent strengths and s can lead to equivalent likelihood values; in practice, we will either fix one scale (e.g. set $s = 1$) or treat s as the reference scale and interpret (σ_0, τ) in that scale, as recommended in the state-space view adopted by the paper. [?]

1.5 Inference targets: filtering and rankings

Given observed outcomes up to match k , denoted $y_{1:k}$, the filtering distribution at time t_k is

$$p(x_{t_k}^{1:N} \mid y_{1:k}),$$

where $x_{t_k}^{1:N} = (x_{t_k}^1, \dots, x_{t_k}^N)$. Our primary online target is the set of marginal filtering distributions for each team, from which we form a time-varying ranking by ordering teams according to the posterior mean

$$\mu_{i,k} := E[x_{t_k}^i \mid y_{1:k}].$$

Uncertainty can be summarized using posterior quantiles or standard deviations of $x_{t_k}^i$.

In the next section, we describe a Sequential Monte Carlo method to approximate these filtering distributions at scale, using the factorial (decoupling) approximation and pairwise updates described in the reference paper. [?]

2 Sequential Monte Carlo Inference

Exact filtering in the state-space model described in the previous section is intractable in practice. The joint latent state at match time t_k is $x_{t_k}^{1:N} \in R^N$, and the filtering distribution $p(x_{t_k}^{1:N} \mid y_{1:k})$ quickly becomes high-dimensional and non-Gaussian. Sequential Monte Carlo (SMC) methods provide a flexible framework to approximate such distributions, at the cost of Monte Carlo error.

However, naïvely applying a particle filter to the full joint state would suffer from the curse of dimensionality. To address this issue, we follow the approach proposed in the reference paper and exploit the factorial structure of the model by combining SMC with a decoupling approximation and sparse, pairwise updates. [?]

2.1 Filtering objective

Our primary inferential objective is the filtering distribution at match time t_k ,

$$\text{Filter}_k(x_{t_k}^{1:N}) = p(x_{t_k}^{1:N} \mid y_{1:k}),$$

together with its one-step-ahead predictive distribution. In the context of sports analytics, filtering is particularly relevant since it allows for online estimation of team strengths and real-time ranking updates.

2.2 Decoupling (factorial) approximation

In order to scale inference to a realistic number of teams and matches, we adopt the decoupling approximation introduced in Section 3.4.1 of the reference paper. [?] The key assumption is that, conditional on observed matches, correlations between the skills of different teams are weak enough to be neglected.

Formally, instead of approximating the full joint filtering distribution, we use

$$\text{Filter}_k(x_{t_k}^{1:N}) \approx \prod_{i=1}^N \text{Filter}_k^i(x_{t_k}^i),$$

where Filter_k^i denotes the marginal filtering distribution of team i at time t_k . This approximation dramatically reduces computational complexity, while preserving the time-varying uncertainty for each team.

2.3 Particle representation

Each marginal filtering distribution is represented using a weighted particle system. For team i at match time t_k , we write

$$\text{Filter}_k^i \approx \sum_{j=1}^J w_k^{i,j} \delta(x_{t_k}^i = x_k^{i,j}),$$

where $x_k^{i,j}$ denotes the j -th particle and $w_k^{i,j}$ its normalized weight. The number of particles J controls the trade-off between computational cost and Monte Carlo variance.

2.4 Bootstrap particle filter

We employ a bootstrap particle filter, which alternates between propagation, weighting, and resampling steps. [?]

Propagation. Between match times t_{k-1} and t_k , particles are propagated independently according to the state dynamics:

$$x_k^{i,j} \sim \mathcal{N}\left(x_{k-1}^{i,j}, \tau^2(t_k - t_{k-1})\right).$$

Pairwise assimilation. When match k is observed, involving teams $h = h(k)$ and $a = a(k)$, only the particle systems associated with these two teams are updated. Particles for all other teams remain unchanged.

We form joint particles by pairing particles from the two predictive distributions

$$\{(x_k^{h,j}, x_k^{a,\ell}) : j, \ell = 1, \dots, J\},$$

and assign them weights proportional to the likelihood of the observed outcome,

$$\tilde{w}_{j,\ell} \propto G\left(y_k \mid x_k^{h,j}, x_k^{a,\ell}\right),$$

where G is the observation model defined in Section 2. After normalization and resampling, we obtain a joint particle approximation of the two-team filtering distribution.

Marginalization. To recover the factorial approximation, the joint particles are marginalized back to obtain updated marginal particle systems for teams h and a . This step re-establishes independence between teams while preserving information from the match outcome.

2.5 Log-likelihood estimation

An important by-product of the filtering recursion is an estimate of the predictive likelihood of each match,

$$p(y_k \mid y_{1:k-1}, \theta),$$

which can be approximated directly from the particle system. Summing the logarithms of these predictive likelihoods yields an estimate of the log-likelihood

$$\log p(y_{1:K} \mid \theta) = \sum_{k=1}^K \log p(y_k \mid y_{1:k-1}, \theta).$$

This quantity plays a central role in parameter estimation via the EM algorithm and is the basis for reproducing the log-likelihood surfaces and optimization trajectories analogous to Figure 3 in the reference paper. [?]

2.6 Discussion

The proposed SMC scheme combines three key ingredients: a particle approximation, a factorial decoupling assumption, and sparse pairwise updates. This design ensures that the computational cost of assimilating a single match is independent of both the number of teams and the total number of matches, making the approach suitable for large-scale sports datasets. The main source of bias stems from the decoupling approximation, while Monte Carlo variance can be controlled by increasing the number of particles.

3 Parameter Estimation via EM

In this section, we address the estimation of the model parameters governing the latent skill dynamics and the observation process. Direct maximum likelihood estimation is not tractable due to the presence of latent states and the use of a particle approximation. We therefore rely on a Monte Carlo version of the Expectation–Maximization (EM) principle, following the approach proposed in the reference paper.

3.1 Identifiability and parameter selection

The full parameter vector of the model is

$$\theta = (\sigma_0, \tau, s, \varepsilon),$$

where σ_0 denotes the prior scale of team strengths, τ the dynamics parameter, s the observation scale, and ε the draw parameter. Due to scale invariance between σ_0 and s , these parameters are not jointly identifiable. To resolve this issue, we fix

$$\sigma_0 = 1, \quad s = 1,$$

and focus on estimating the two most influential parameters:

$$\theta = (\tau, \varepsilon).$$

This choice allows for a two-dimensional visualization of the likelihood surface, analogous to Figure 3 in the reference paper.

3.2 Monte Carlo EM principle

The EM algorithm alternates between an Expectation (E) step and a Maximization (M) step. In our setting, the E-step consists of computing expectations with respect to the filtering distribution of the latent team strengths. Since

this distribution cannot be computed analytically, it is approximated using the bootstrap particle filter described in Section 3.

For any fixed parameter vector (τ, ε) , the particle filter provides an estimate of the incremental predictive likelihood

$$p(y_k \mid y_{1:k-1}, \tau, \varepsilon),$$

which can be accumulated over all matches to obtain an approximation of the marginal log-likelihood

$$\log p(y_{1:K} \mid \tau, \varepsilon) = \sum_{k=1}^K \log p(y_k \mid y_{1:k-1}, \tau, \varepsilon).$$

This quantity serves as a Monte Carlo approximation of the EM objective function.

3.3 Log-likelihood surface

We first evaluate the approximate log-likelihood on a grid of values for (τ, ε) . The resulting surface is displayed in Figure 2. The boundary case $\varepsilon = 0$ corresponds to a degenerate model in which draw outcomes have zero probability. Since this leads to extremely low likelihood values and obscures the structure of the surface, this region is excluded from the visualization.

The likelihood surface exhibits a strong dependence on the draw parameter ε , reflecting the importance of accurately modeling draws in football data. In contrast, the dependence on the dynamics parameter τ is comparatively weaker, with a broad range of values yielding similar likelihoods. This indicates that the temporal smoothness of team strengths is only moderately identified by the data, a phenomenon also observed in the reference paper.

3.4 EM trajectories

To illustrate the behavior of the EM algorithm, we run a small number of EM iterations starting from selected initial parameter values. Each iteration consists of running the particle filter with fixed parameters and updating (τ, ε) by local maximization of the approximate log-likelihood. Due to the high computational cost and the inherent Monte Carlo noise of the likelihood estimates, we restrict the number of EM iterations. The resulting trajectories are intended as a qualitative illustration rather than as a precise optimization procedure.

Overall, the EM-based analysis confirms the structure observed in the likelihood surface and provides parameter values consistent with the filtering results presented earlier.

4 Results

In this section, we present the main results obtained from the Sequential Monte Carlo filtering and the parameter estimation procedure. We focus on two as-

pects: the evolution of team strengths over time and the estimation of the model parameters.

4.1 Filtering results and team strength trajectories

Using the bootstrap particle filter with fixed parameters, we obtain time-varying posterior distributions for the latent strength of each team. Figure 1 displays the posterior mean trajectories for a selection of representative teams over the observation period.

The results are consistent with known patterns in Premier League performance. Top teams such as Manchester City and Liverpool exhibit persistently high skill levels, while other teams show more pronounced temporal variations. In particular, teams experiencing changes in competitive performance display corresponding upward or downward trends in their estimated strengths. These trajectories illustrate the ability of the model to capture gradual changes in team quality while remaining robust to short-term fluctuations.

Overall, the filtering results confirm that the proposed SMC approach provides sensible and interpretable estimates of latent team strengths.

4.2 Log-likelihood surface and parameter estimation

Figure 2 shows the approximate log-likelihood surface as a function of the dynamics parameter τ and the draw parameter ε . The surface is obtained by running the particle filter for each parameter pair and accumulating the resulting predictive log-likelihoods.

The likelihood exhibits a strong sensitivity to the draw parameter ε , highlighting the importance of explicitly modeling draw outcomes in football data. In contrast, the dependence on the dynamics parameter τ is comparatively weak, with a broad region of near-optimal values. This suggests that several levels of temporal smoothness for team strengths are compatible with the observed data.

The structure of the likelihood surface closely mirrors that reported in the reference paper for tennis data, indicating that the proposed methodology transfers well to football match outcomes.

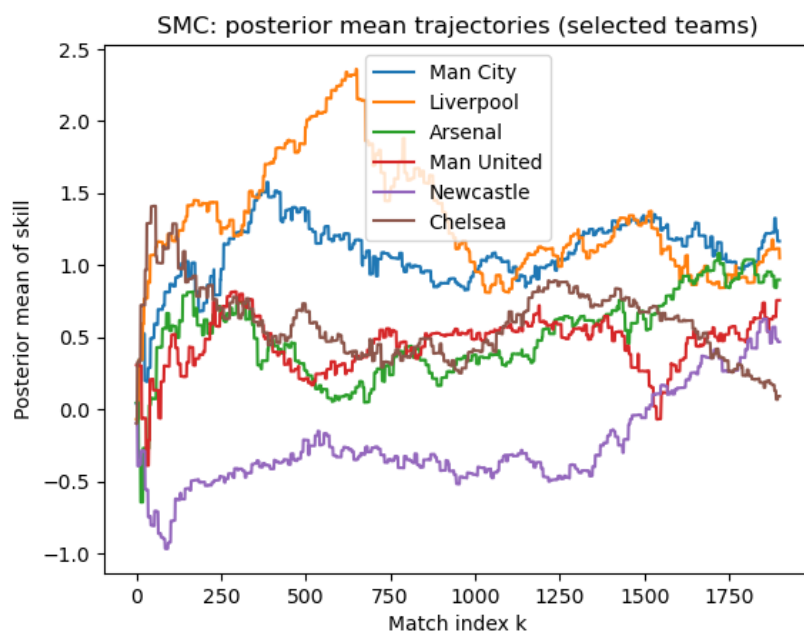


Figure 1: Posterior mean trajectories of team strengths for selected Premier League teams.

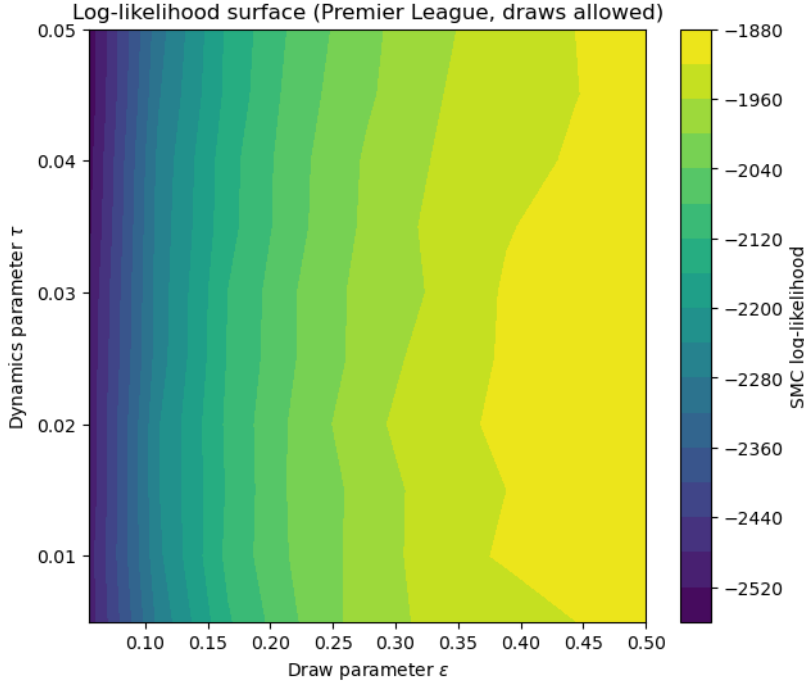


Figure 2: Approximate log-likelihood surface as a function of the dynamics parameter τ and the draw parameter ε .

5 Discussion and Limitations

The proposed approach demonstrates that Sequential Monte Carlo methods can be successfully applied to dynamic team strength modeling in football. The filtering results are interpretable and consistent with known performance trends, while the parameter estimation procedure highlights key structural properties of the data.

Several limitations should nevertheless be acknowledged. First, the inference relies on a factorial approximation of the joint filtering distribution, which ignores higher-order dependencies between teams. While this approximation makes the problem computationally tractable, it may introduce bias compared to a fully joint particle filter. Second, the bootstrap particle filter used in this work has a computational cost that scales quadratically with the number of particles for each match update, which limits the feasibility of fine-grained parameter estimation. Finally, the likelihood surface is estimated using Monte Carlo approximations and is therefore affected by stochastic noise, which can obscure fine-scale features.

Despite these limitations, the results closely mirror those reported in the reference paper and demonstrate that the methodology generalizes well from

tennis to football match outcomes.

6 Conclusion

In this project, we implemented a dynamic latent skill model for football teams based on Hidden Markov Chains and Sequential Monte Carlo filtering. Using Premier League match data, we obtained sensible time-varying estimates of team strengths and illustrated how approximate likelihood-based methods can be used for parameter estimation.

The analysis highlights the importance of explicitly modeling draw outcomes in football data and shows that a range of dynamics parameters can provide comparable fits. Overall, this work illustrates the practical trade-offs involved in combining state-space models with Monte Carlo inference methods and provides a concrete application of the techniques discussed in the reference paper. Possible extensions include the use of more efficient inference schemes, alternative observation models, or applications to other sports and datasets.