Imperial College London

Final Report

Design of computationally efficient algorithm for predicting the evolution of eczema severity

Samuel Oliveira

Supervisors: Dr Reiko Tanaka and Ariane Duverdier

June 2023

Submitted in partial fulfilment of the requirements for the award of MEng in Biomedical Engineering from Imperial College London

Word Count: X words

Contents

1 Introduction **2**

2 Methods **4**

2.1 Dataset 4

2.2 Model overview 5

2.3 Model validation 8

2.4 Choice of method 8

2.4.1 Kalman Filter 9

2.4.2 Particle Filter 10

2.4.3 Variational Inference 13

2.4.4 Comparison between methods 14

2.5 Particle Filter 15

2.6 Parameter Tuning 15

3 Results and Discussion **16**

3.1 Parameter tuning 16

3.1.1 Extent 16

3.1.2 Intensity signs 18

3.1.3 Subjective signs 19

3.1.4 Number of particles 19

3.1.5 Parameter tuning limitations 20

3.2 Prediction of severity items 20

3.3 PO-SCORAD prediction 23

3.4 Computational efficiency 24

3.5 Future steps 26

4 Conclusion **27**

A EczemaPF models **31**

A.1 Extent models 31

A.2 Intensity models 31

A.3 Subjective models 32

B Supplementary results **34**

B.1 Intensity signs 34

B.2 Number of particles 35

Abstract

Eczema (atopic dermatitis, or AD) is a chronic skin disease, for which designing personalised treatment strategies, as opposed to "one-size-fits-all" approaches, is of high clinical relevance. Designing personalised treatment strategies requires an accurate prediction of the evolution of eczema severity because disease symptoms fluctuate dynamically every day and manifest as relapses and remissions. Past work has used offline learning methods such as Markov-Chain Monte Carlo to predict the evolution of eczema severity scores. However, these methods perform offline learning, with the time required for prediction growing with the number of days in the time-series data. For a time-series with about 42 days, past work takes approximately 40 seconds per patient in the dataset to output a patient's daily prediction. This number grows without a bound as the number of days in the time-series increases, as well as the number of patients in the dataset. Thus, they are unsuitable for near real-time prediction that can guide patients towards daily treatment. To address this problem, I propose an online inference framework that can output predictions in a matter of seconds. This framework uses Sequential Monte-Carlo methods to avoid some of the pitfalls of offline learning. The framework is used in a state estimation setting, and achieves a predictive performance comparable to past work despite simpler non-hierarchical models. The daily training process for each patient is reduced to 5 seconds per patient in the dataset, and is now independent of the number of days the framework has been used. This reduced training time means, in the future, this framework could be implemented as a severity management tool, providing patients with a prognosis of their eczema, as well as treatment recommendations to guide them towards the appropriate care required to manage their eczema severity.

1 Introduction

Eczema (also known as atopic dermatitis, or AD) is a chronic inflammatory skin disease affecting 20% of children worldwide [1]. Common symptoms, such as dry and itchy skin, are highly unpredictable, manifesting as periods of flares and remission without a clear pattern. Historically, treatment strategies have relied on the application of emollients or anti-inflammatory creams. Predicting the effects of treatments for a specific patient is challenging due to the sudden fluctuations in symptoms that most patients showcase. This is exacerbated by the fact that the patient's response to the same treatment can change for different time points.

Furthermore, both symptom progression and treatment response vary greatly between patients [2]. It is thus imperative to introduce personalized treatment regimes that take into account not only the dynamics of a specific patient's symptoms, but also how effective a treatment has proved to be in the past.

Eczema severity is commonly reported as a single scalar score, obtained by aggregating the scores for multiple severity items corresponding to different characteristic eczema symptoms. Each severity item score aims to capture a different eczema symptom with its own specific dynamics. The symptoms considered, as well as their respective contributions to the eczema score, vary between different measures. SCORing AD (SCORAD), as well as Eczema Area and Severity Index (EASI) have both been validated as outcome instruments for the measurement of eczema severity in clinical trials [3, 4, 5].

In 2022, the Tanaka Group introduced EczemaPred [6], a computational framework aimed at predicting the evolution of eczema severity at a patient level using Bayesian state-space models. This framework models and predicts each severity item separately, and aggregates these predictions to obtain a predictive eczema severity score. This approach allows not only for the extraction of more information from the data, but also reduces the uncertainty in the prediction of the aggregated score. Furthermore, by providing probability distributions instead of point predictions of the scores, the Bayesian approach allows for better interpretability of the uncertainty associated with each prediction. The Bayesian state-space framework assumes each recorded severity item results from imperfect observations of a true latent severity, with the latter following its own dynamics. Despite being a general framework which can be applied to any eczema severity score, its results were validated while predicting Patient-Oriented SCORAD

[7, 8] (PO-SCORAD) , a self-assessed version of SCORAD that can be recorded by patients themselves on a smartphone app. EczemaPred was successful in predicting patient-specific eczema severity, outperforming standard time-series forecasting models such as auto-regressive models.

In the future, a similar framework could be integrated into a smartphone app, alongside deep learning systems that assess eczema severity from camera images [9], to allow patients to track the daily evolution of their eczema symptoms, as well as receive personalised predictions for the days ahead and recommended treatment regimes. The integration into a smartphone app requires prediction to be obtained in just a few seconds. However, the current version of EczemaPred takes up to half an hour to train on a full dataset, and thus to output a prediction.

The main reason behind this computational inefficiency is the use of Hamiltonian Monte Carlo to perform inference. Similarly to other Markov Chain Monte-Carlo (MCMC) algorithms, it only supports offline learning. Thus, each model has to be re-trained on all datapoints (which takes up to half an hour) every time a new measurement is obtained. On the other hand, online methods prove to be significantly more efficient for time-series data, as only the latest point is used to train the model at each time-point. It is thus critical for EczemaPred to support online learning, as it would allow for a drastic reduction in the time taken to output a prediction every day, and thus enable EczemaPred's integration into real-world scenarios.

This project aimed to improve the computational efficiency of EczemaPred. Several techniques were explored, including Kalman filters, Sequential Monte-Carlo, and Variational Inference. The methods were reviewed to assess their suitability for the problem at hand, taking into account the additional challenges presented by eczema severity data. Firstly, each severity sign will be a discrete ordinal outcome, with rapid and large changes in its value (corresponding to exacerbations of disease symptoms). Secondly, these dynamics vary significantly between patients, and thus the chosen method should allow for a patient-specific model. Finally, eczema severity datasets, including the clinical dataset PFDC [10] utilized in this project, have significant proportions of missing values. If daily recordings are expected, the PFDC dataset has 13.6% missing values [6]. Thus, the chosen method should allow for missing data. Following the choice of method, the chosen method ought to be implemented assuming the same latent dynamics and measurement distributions as EczemaPred (where possible), fit on the PFDC dataset, and its predictive performance and computational efficiency compared with

EczemaPred's.

In this report, I introduce EczemaPredFast (EczemaPF), an EczemaPred-inspired framework to predict the dynamic evolution of eczema severity in patients. Rather than utilizing MCMC to perform inference, this version of EczemaPred employs Sequential Monte-Carlo, an online approach to state and parameter inference also commonly known as Particle Filter (PF). The framework consists of a collection of Bayesian state-space models, describing the dynamics of each severity item. The models for each severity sign are mimicked from EczemaPred where possible, and adapted where necessary. The framework is used to predict PO-SCORAD using a longitudinal clinical study dataset [10], and its predictive performance and computational efficiency are compared to the original EczemaPred's on the same dataset.

2 Methods

2.1 Dataset

The PFDC dataset [10] was used to train the models. This dataset was obtained from an observational clinical study. The data consists of daily PO-SCORAD scores (and its breakdown for each severity sign) recorded by 16 adult patients for up to 12 weeks, with a total of 1136 observations and 13.6% missing values. Patients recruited had SCORAD scores between 20 and 40 (mean SCORAD 34.6, SD=4.4 at the start of the study). Patients used an app to record PO-SCORAD daily while continuing their treatment. Eight of the 16 patients presented AD flares during the study, and their medication was changed accordingly.

PO-SCORAD can be calculated using the formula 0.2 A + 3.5 B + C. A is the extent (the percentage of skin area affected by eczema in the patient's whole body). B corresponds to intensity signs, and is the sum of the scores for six intensity signs (dryness, redness, swelling, oozing, scratching and thickening), each expressed on an ordinal scale as 0 (absent), 1(mild), 2 (moderate) and 3 (severe). C corresponds to subjective signs, and is the sum of two signs (itching and sleep disturbance), each assessed in a visual scale from 0 (no symptom) to 10 (severe symptom).

All signs take discrete values, with extent and intensity signs having a resolution of 1, whereas

subjective signs have a resolution of 0.1. Two example trajectories of PO-SCORAD dynamics (and the breakdown for each severity sign) for two patients in the PFDC dataset is shown in Figure 1.

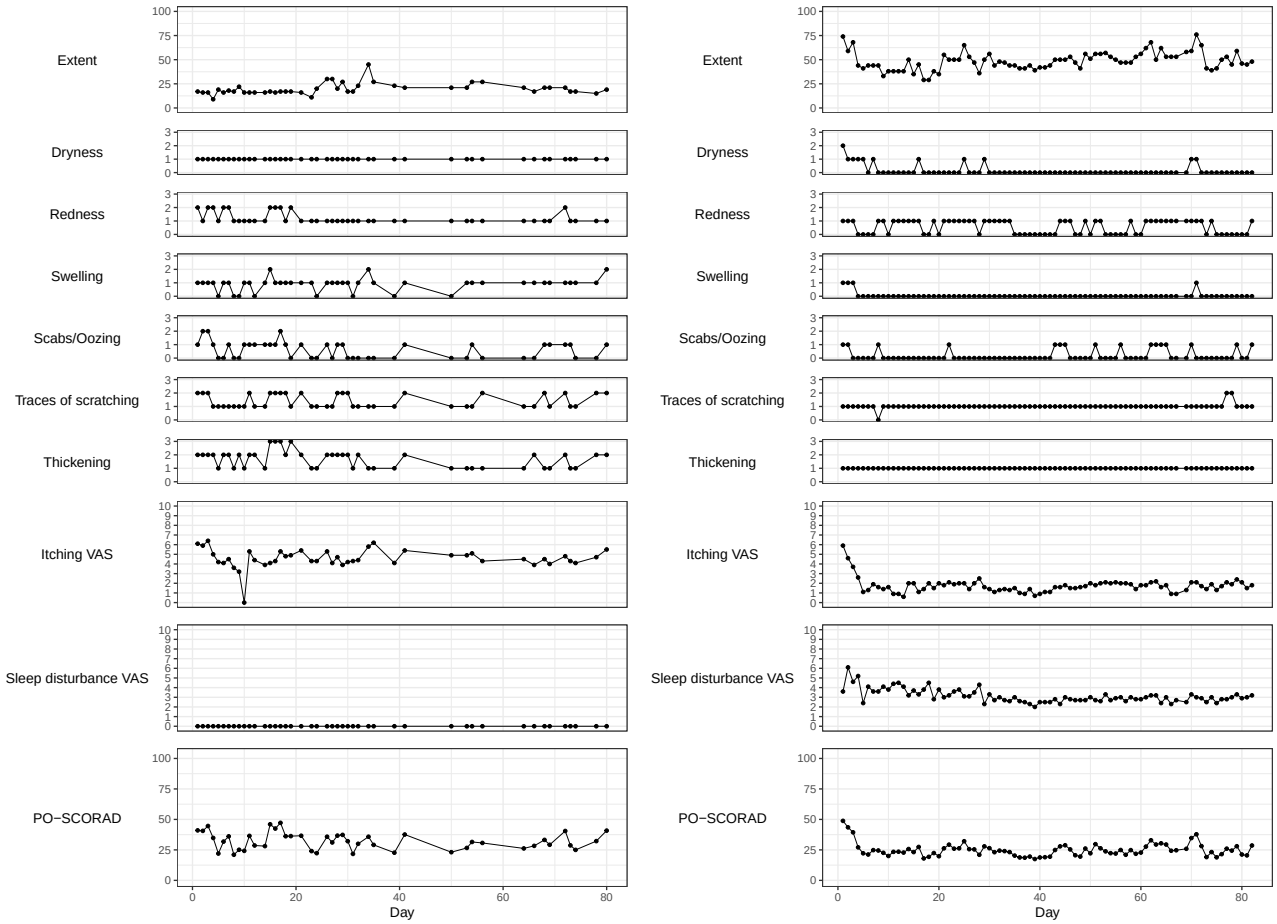


Figure 1: Example trajectories of PO-SCORAD and the respective severity items necessary to calculate it for two different patients in the PFDC dataset.

2.2 Model overview

EczemaPredFast is a set of multiple Bayesian state-space models, each used to describe the data-generating mechanism of a specific severity item. Each model assumes the existence of a true latent/unobserved severity, following its own dynamics. Each observation by a patient (i.e. each datapoint) is the result of an imperfect measurement of the latent severity, as shown in Figure 2.

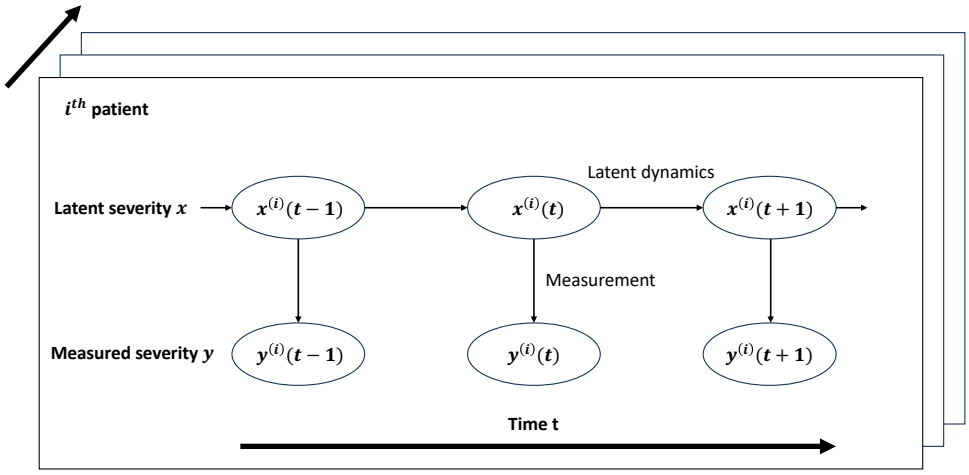


Figure 2: Overview of Bayesian state-space models, as utilized in both EczemaPF and EczemaPred [6].

EczemaPredFast is used in this report to predict PO-SCORAD, and thus contains nine models, one for each severity item necessary to calculate PO-SCORAD. However, this framework can easily be adapted to predict different eczema severity scores. The individual predictions of each model were aggregated to predict PO-SCORAD, assuming thus the severity items evolve independently, as shown in Figure 3.

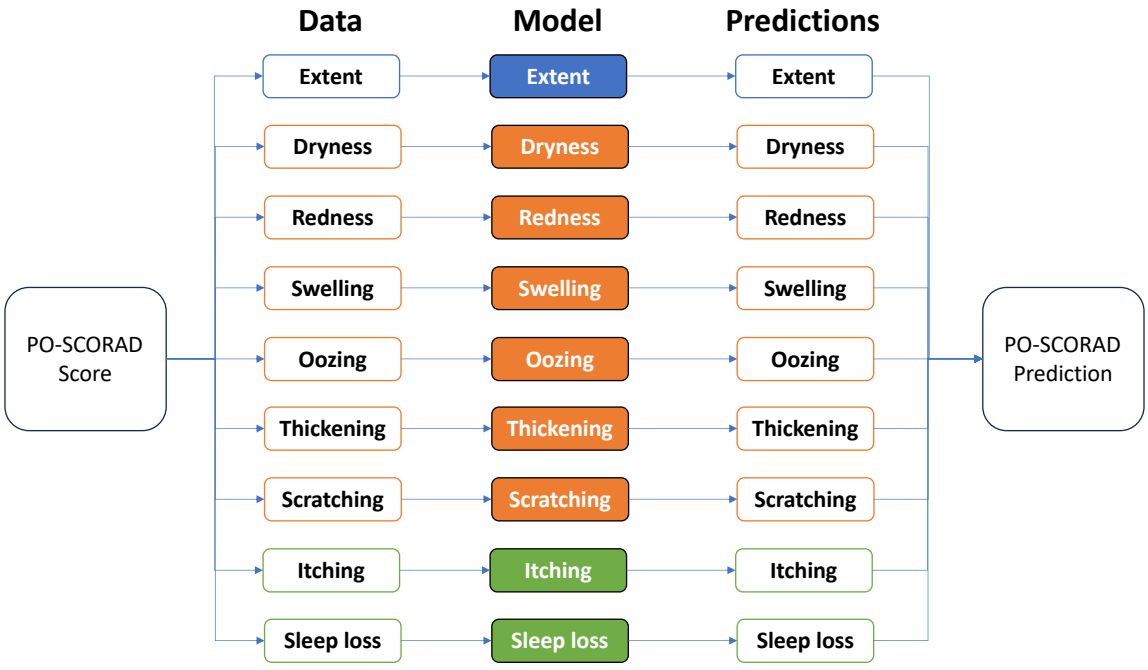


Figure 3: Framework overview for both EczemaPF and EczemaPred [6].

The chosen latent dynamics and measurement distributions choices for EczemaPF and EczemaPred

are shown in Figure 3. The latent dynamics were all modelled as Gaussian random walks, in the absence of clear insights into the biological mechanisms behind the evolution of each severity sign. As shown in section 3.1.1., EczemaPF diverged for Markov Chain latent dynamics, and thus a random walk was implemented. The measurement was modelled as a binomial distribution. For the intensity signs, both the latent and measurement dynamics were modelled as Gaussian random walks. The latent gaussian random walk mirrors EczemaPred's latent intensity sign dynamics, whereas a Gaussian random walk is used as observation distribution due to difficulties in implement an ordinal logistic distribution (as in EczemaPred). Finally, for the subjective signs, a latent variable followed a Gaussian random walk, with its observation being described by a binomial distribution, as is also the case with extent.

EczemaPred

Severity component	Latent dynamics	Measurement distribution
Extent	Markov Chain	Binomial
Intensity signs	Random walk	Ordered logistic
Subjective symptoms	Random walk	Binomial

EczemaPF

Severity component	Latent dynamics	Measurement distribution
Extent	Random walk	Binomial
Intensity signs	Random walk	Random walk
Subjective symptoms	Random walk	Binomial

Figure 4: Latent dynamics and measurement distributions for the three types of severity items used to calculate PO-SCORAD, for both EczemaPred and EczemaPF.

Parameters for these models (latent and observation models) were set (as described in section 2.6), instead of inferred. Thus, the particle filter for each sign used two models with fixed parameters (the same for all patients) to predict future observations. This differs from EczemaPred, where the framework inferred patient-specific parameters, and used these to predict future observations. Priors for latent states were also made patient-independent, and weakly informative. This is due to the large degree of uncertainty in the values of the latent severity states at the beginning of the trials. Missing values were treated by the particle filters as an absence of measurement. Details of these models, priors and choices of parameters are given in Appendix A.

2.3 Model validation

The predictive performance on EczemaPF was evaluated in forward-chaining setting, as described in Figure 5. This means each model's performance was evaluated at every 4 days of training. Thus, the model is first trained on the first day, and its 4-day-ahead predictive performance is assessed. Afterwards, the model is re-trained on the first 5 days of data, and its performance assessed on 4-day-ahead forecasts. This process is repeated until the end of the time-series. Figure 5 shows a schematic of this training process, for EczemaPred (left) and EczemaPF (right). While the predictive performance results will not be affected by any differences, the use of a cache for online methods such as EczemaPF results in the computational efficiency observed in section 3.4.

The Bayesian approach means the framework outputs a prediction represented by a probability distribution instead of a single estimate. Thus, we chose a logarithmic scoring rule, the log predictive density (lpd) [11]. A higher lpd is associated with more accurate and certain predictions. Both the individual severity item predictions, as well as the aggregated PO-SCORAD predictions are evaluated. The lpd was plotted as a function of the number of days (related to the number of training data points), resulting in learning curves that can be compared with EczemaPred's.

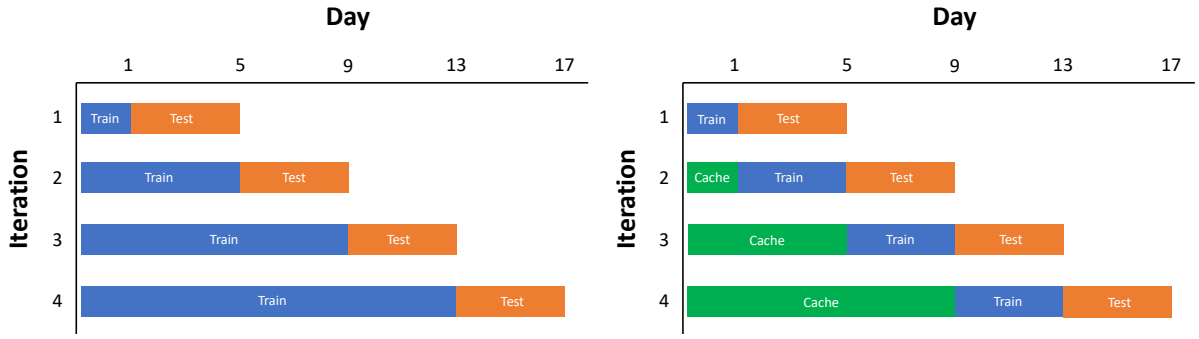


Figure 5: Comparison of the forward-chaining training setting for offline mehtods (left) and online methods (right).

2.4 Choice of method

The first main objective of this project was to explore techniques with the potential to increase the computational efficiency of EczemaPred, with a focus on those suitable for online learning.

A thorough literature review was performed, with a summary of it present in Section 3 of my Planning Report. A short discussion on the advantages and drawbacks of each technique is given in this section. Kalman filters, Particle filters and Variational Inference were shortlisted as possible methods.

2.4.1 Kalman Filter

The Kalman filter [10] is an algorithm suitable for state-space models, aimed at estimating hidden variables in state-space models. This filter provides an optimal solution when the model dynamics are linear, and each hidden variable's estimate can be described as a Gaussian distribution. The filter propagates a Gaussian representation (defined by a mean value $\hat{x}_k$ and a covariance matrix P_k) of the hidden state in time, using linear dynamics. At each time-point, the filter firstly predicts the latent state at the next time-step, and then corrects this prediction based on the new data point received. Thus, the filter uses the transition dynamics model to guess how the latent variable will evolve in time (prior distribution), and then corrects this prediction based on the data that has arrived at the latest data point (obtaining the posterior distribution). Figure 6 shows a schematic of the three steps in each iteration of the Kalman filter. The notation in Table 1 is followed.

Table 1: Notation for Kalman Filter diagram.

Symbol	Meaning
k	Time
$\hat{x}_k$	Estimated Latent Vector
P_k	Covariance Matrix
z_k	Measurement Vector
u_k	Input Vector
$\hat{x}_{k,k-1}$	Predicted Latent Vector
$P_{k,k-1}$	Predicted Covariance Matrix
A	Transition Matrix
B	Control Matrix
H	Observation Matrix
Q	Process Noise Matrix
R	Measurement Noise Matrix

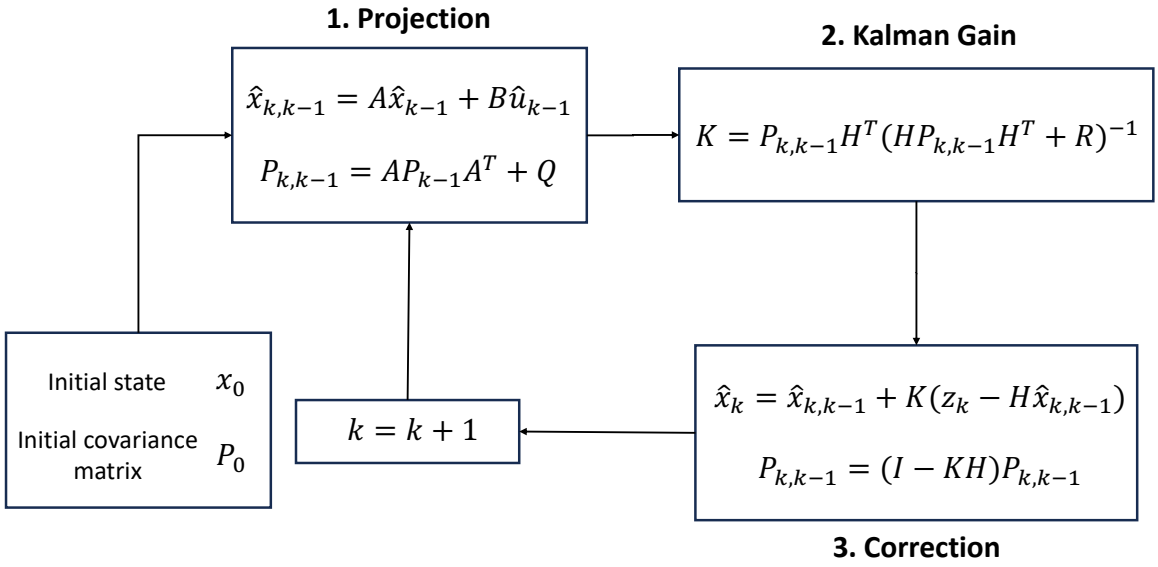


Figure 6: Schematic of each iteration of the Kalman filter, composed of three stages: projection, calculation of the Kalman Gain, and correction.

2.4.2 Particle Filter

Sequential Monte Carlo methods are based on the idea of representing any probability density function by a set of samples called particles. When used to estimate the latest latent variable (filtering operation), these are commonly known as Particle Filters (PF). The filter aims to represent the latent variable at each time point as a probability distribution. Each particle represents a guess for the true latent variable, and the set of particle creates a discrete approximation of any probability distribution. Thus, with enough particles, the filter can in theory represent any distribution, no matter its complexity.

Each iteration of the simplest form of this algorithm, the bootstrap filter [12], can be divided into three steps, as shown in Figure 7. The notation in Table 2 will be followed.

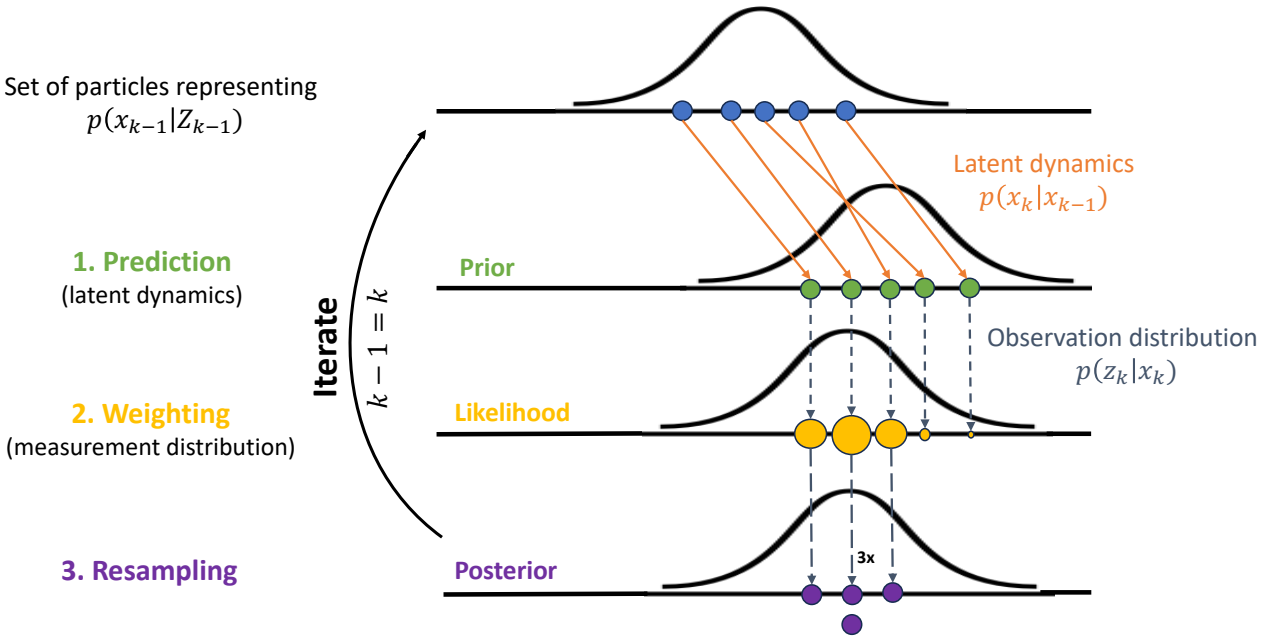


Figure 7: Schematic of each iteration of the bootstrap filter [12], composed of prediction, weighting of particles, and resampling steps [13].

Table 2: Notation for Particle filter equations.

Symbol	Meaning
k	time
X_k	$\{x_j, j = 0, ..., k\}$.
x_k	Latent Variable at time k.
Z_k	$\{z_j, j = 0, ..., k\}$.
z_k	Measurement at time k.
x_k^i	Latent Variable of particle i at time k.
w_k^i	Weight of particle i at time k.

1st Transition

Let us have a set of uniformly weighted particles at time k-1, each with a specific value of x, representing the distribution $p(x_{k-1}|Z_{k-1})$. We want to predict the latent variable at the next time step k. That is, we want to find $p(x_k|Z_k)$, equivalent to transitioning the particles to the next time step. In a Monte-Carlo framework, samples from this distribution would approximate it. However, as this is the distribution we are trying to find, we cannot sample from it. Thus, a mathematical trick, Importance Sampling, is introduced [14]. This trick shows that, under specific conditions that can be assured in the running of the Particle Filter, samples from a different distribution q can accurately approximate samples from our distribution of interest p,

if we weigh them accordingly. This is achieved by sampling from distribution q, and multiplying these samples by the ratio of these distributions. This can be described by:

$$E_p[f(x)] = E_q[\frac{p(x)}{q(x)}f(x)]$$

The weighted set of samples will approximate the distribution of interest p, even if the samples were not originally obtained from this distribution. How to choose the distribution q to sample from to approximate p will become apparent in the next section.

2nd Weighting of particles

It can be shown [15, 16, 17] that the weights of each particle i at time k can be described by:

$$w_k^i \propto w_{k-1}^i \frac{p(z_k|x_k^i)p(x_k^i|x_{k-1}^i)}{q(x_k^i|x_{k-1}^i, z_k)}$$

We can observe that if we choose the importance density $q(x_k|x_{k-1}^i, z_k)$ to be the transition probability $p(x_k^i|x_{k-1}^i)$, then the previous equation simplifies to:

$$w_k^i \propto w_{k-1}^i p(z_k|x_k^i)$$

Thus, starting with a set of particles that approximates $p(x_{k-1}|Z_{k-1})$, by choosing the importance density q in the 1st step to be the transition probability $p(x_k^i|x_{k-1}^i)$, and then weighing the particles by the likelihood distribution $p(z_k|x_k^i)$, we obtain a set of particles that approximates the posterior $p(x_k|Z_k)$. The filter thus tracks the latent variable by using its knowledge of the transition model, and the data at time k.

3rd Resampling

The first two steps are enough for the filter to track the hidden latent variable. However, due to the heavily documented problem of particle degeneracy [18, 19], a third step is added. Particle degeneracy happens due to the weighting step, in which particles that explain the data well (having high values of $p(z_k|x_k^i)$) get large weights, and other particles get small weights.

Over time, this means only a very small number of particles will have meaningful weights, and computational resources will be wasted on particles with small weights that do not make a significant contribution to the posterior $p(x_k|Z_k)$. This will limit the ability of the filter to express any probability distribution, as it will have a very limited number of particles that actually determine this distribution.

Thus, in a 3rd step in each iteration, N new particles are randomly selected, with replacement, from the previous set of N weighted particles, with a probability proportional to its weight.

These three steps are repeated at each iteration at time k, keeping track of the latent variable x as each new data point arrives. Resampling has a destructive effect on the diversity of the values of particles in the set [18, 20]. This can lead to a phenomenon called sample impoverishment, where the set of particles where most particles have significant weightings (unlike particle degeneracy), but most particles represent the same (or very similar) latent variable value. Thus, resampling should only occur at certain iterations. The effective sample size is the main criteria used to decide when to resample [18], as it represents the proportion of particles with a non-neglibile weight in the set. When this number drops below a certain threshold, particle degeneracy is starting to occur, and resampling is performed to prevent it.

2.4.3 Variational Inference

Variational Inference is a method based on approximating a distribution using a combination of simpler distributions. In this context, simpler distributions would be combined to approximate the posterior $p(x_k|Z_k)$. Firstly, a family of distributions defined by a set of parameters has to be selected, for example a family of Gaussians where their means and variances can be varied. Variational Inference then optimizes these parameters to find the best approximation to the true posterior $p(x_k|Z_k)$. This is done by minimizing the Kullback-Leiber divergence between the approximate and trus posteriors. Thus, the algorithm is an optimization procedure, which can be performed with techniques such as stochastic gradient descent, making it particularly powerful in large-scale problems [21].

2.4.4 Comparison between methods

Both the Kalman filter and the particle filter are online methods, only requiring the latest data point to update their predictions of the latent variable. The Kalman filter's biggest strength is its simplicity, both conceptually and computationally. Its time complexity is of $O(n^{2.376})$[22], with n being the dimension of the state space. In this use case, that corresponds to the number of severity signs used to calculate PO-SCORAD, and thus $n = 9$, and the Kalman filter provides a computationally efficient solution. However, the filter's optimality relies on the assumptions of linear dynamics and Gaussian states. As these are likely to be violated for the highly unpredictable eczema severity data, it is likely the very simple solution of the filter would be suboptimal.

Variational Inference can be made online or offline depending on the choice of optimization procedure. As one of the main objectives of this project is to find an online method (meaning its time complexity does not scale with the number of training time points), we focus specifically on Variational Inference using stochastic gradient descent (SGD), which allows for an online implementation. However, for time-series data, consecutive samples have high degrees of correlation, which leads to unstable training when using SGD [23]. Using mini-batches can partly solve this issue, but it will increase the computational complexity of the algorithm [23]. Furthermore, the method relies on the choice of a family of distributions to create the approximation to the true posterior. There is a lack of guidance on how to choose the most appropriate family, in particular for eczema severity prediction.

The Particle filter is a fully online method that does not rely on the assumptions of the Kalman filter. It can capture more complex distributions and non-linear model dynamics, which are likely to lead to better accuracy than the latter. Furthermore, it supports an online implementation better than Variational Inference, and does not require the choice of a family of distributions. The main issue with this method is the possibility of sample impoverishment [18, 20]. However, this issue can be avoided if the latent dynamics and measurement distributions chosen have significant process noise. Due to the high degrees of uncertainty in latent eczema severity dynamics, the latent variable dynamics will be highly variable, and thus particle degeneracy is expected to be avoided. However, results were analysed to confirm this was the case, and that the filter did not diverge during training.

It was concluded that for the timespan of this project, the Particle filter provided the best compromise between speed, accuracy and ease of implementation. Crucially, the particle filter presents multiple ways to reduce its time complexity, such as hyperparameter tuning and straightforward parallel implementations. Furthermore, its main issue of sample impoverishment is likely to be circumvented.

2.5 Particle Filter

Model inference was performed using a Particle filter algorithm for each severity sign. At each time point, the filter updated its set of particles representing its guess of the latent severity item, and forecasted the observed severity item 4 days ahead.

EczemaPF was created using bootstrap Particle filters, as implemented by the pypfilt Python package [1], which has previously been used to forecast infectious disease epidemics [24, 25]. Each filter was run with 1000 particles.

Resampling was performed when the effective sample size dropped below a threshold of 0.5. This value was set higher than usual [18] due to the low number of particles used (1000). Finally, an unexplainable drop in performance as the time in the time-series data increases is characteristic of sample impoverishment, showing the filter has diverged. Thus, sample impoverishment was checked by visual observation of the learning curves of the filter.

2.6 Parameter Tuning

The results presented in the following section were the result of a parameter tuning process. For each type of sign (extent, intensity or subjective), the values of parameters were chosen based on which ones had the highest lpd out of the combinations of parameters tested. As the different severity items that compose PO-SCORAD are considered to be independent of each other by the models used, this analysis did not have to take into account any dependencies.

The Particle filter is used in a state-estimation setting, and thus the filter has to assume the parameters for the latent dynamics and observation distributions of each severity item.

[1] https://pypi.org/project/pypfilt/

EczemaPred, on the other hand, infers these using the data given, while also tracking the latent severity as EczemaPF does.

For extent, the latent variable was described by a Markov chain, as in EczemaPred. As the filter diverged (section 3.1.1.), a Gaussian random walk was used instead, due to the absence of clear insights into the biological mechanisms behind its evolution. The standard deviation was tuned. For intensity signs, an ordinal logistic distribution could not be used as an observation model due to implementation difficulties, and thus a Gaussian random walk was used both for the latent and observation models. The standard deviations were tuned. For subjective signs, the standard deviation of the latent Gaussian random walk was tuned. Finally, the number of particles needed was varied, and its impact on performance assessed.

3 Results and Discussion

All EczemaPF models were fitted successfully on all severity signs data. An informal analysis of potential divergence was performed by visual checks of the learning curves of each particle filter. Section 3.1. shows the results of the parameter tuning process, whereas sections 3.2. and 3.3. show the results with the optimal set of parameters found.

3.1 Parameter tuning

3.1.1 Extent

Markov Chain

The latent dynamics for the extent are modelled as a Markov chain for EczemaPred. EczemaPred learns the parameters of this Markov chain (that is, the transition probabilities and steady state distribution). However, since EczemaPF is used in a state-only estimation setting, these transitions are fixed. Thus, at each step, each particle representing a guess of the latent probability of the extent evolves in a fully deterministic setting. No noise is added to the particle set, and thus we observe divergence of the filter (its performance decreases as it is fed more data), due to sample impoverishment, as seen in Figure 8.

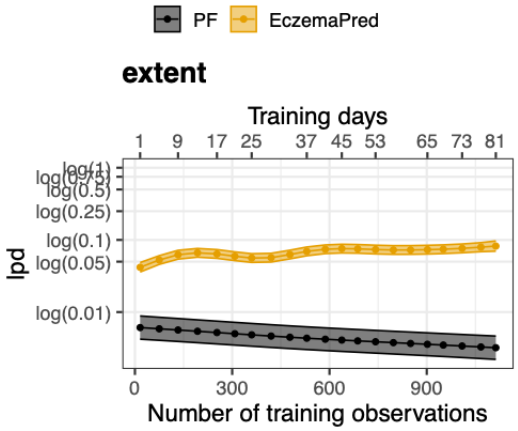


Figure 8: Predictive performance (evaluated using lpd) for 4-day ahead forecasts as a function of the number of training observations, for the Extent model of EczemaPF assuming Markov chain latent dynamics.

Gaussian RW Tuning

Due to the divergence of the filter when using a Markov chain, the latent extent probability was chosen to be a Gaussian random walk, due to the lack of insight into the mechanism behind the dynamics of this severity item. The standard deviation of this Gaussian was varied, as well as the scale in which this Gaussian was set, and its result on lpd at 65 days of training is shown in Table 3.

Table 3: Predictive performance (lpd) for different Extent latent dynamics, with different parameters.

Distribution	**Standard Deviation**	**lpd**
Gaussian RW	0.2	log(0.01)
Gaussian RW	0.05	log(0.04)
Gaussian RW	0.02	log(0.055)
Gaussian RW	0.01	log(0.07)
Gaussian RW	0.005	log(0.067)
Gaussian RW (in logit scale)	0.3	log(0.05)
Gaussian RW (in logit scale)	0.1	log(0.065)
Gaussian RW (in logit scale)	0.05	log(0.058)

A standard deviation of 0.01 (variance of 0.001) was chosen due to the higher value of lpd achieved.

3.1.2 Intensity signs

Intensity signs dynamics are defined by two parameters: the standard deviations of the latent model and the observation model. These parameters were varied and the performance of EczemaPF assessed. Furthermore, the priors on the latent intensity were also varied. The results are presented in Appendix B.1.

Only redness was improved by an increase in both standard deviations, with all other signs performing equally or worse. Further work should focus on the effects on predictive performance of keeping the latent standard deviation constant while altering the observation standard deviation, and vice-versa.

The prior tuning process showed dryness and oozing benefited from a less informative prior, whereas the other signs' performances worsened.

3.1.3 Subjective signs

For the subjective signs, only 1 parameter (the standard deviation of the Gaussian random walk) could be changed. The results are presented in Table 4. It can be seen that the standard deviation changes had little impact on performance. Analysis of the stability of the learning curves suggested 0.3 is the most appropriate standard deviation for both signs, as the learning curves began at larger lpd values and converged, more smoothly, to around the same lpd at 65 days.

Table 4: Predictive performance (lpd) for fixed subjective signs latent dynamics, with different parameters.

Sign	**Distribution**	**Standard Deviation**	**lpd**
Sleep	Gaussian RW (in logit scale)	0.3	log(0.20)
Sleep	Gaussian RW (in logit scale)	0.2	log(0.19)
Sleep	Gaussian RW (in logit scale)	0.05	log(0.195)
Itching	Gaussian RW (in logit scale)	0.3	log(0.033)
Itching	Gaussian RW (in logit scale)	0.2	log(0.035)
Itching	Gaussian RW (in logit scale)	0.05	log(0.0342)

3.1.4 Number of particles

While a larger than necessary number of particles should not have any impact on results, a smaller than needed number results in worse results. Thus, the previous experiments were performed for a high number of particles (5000). This was to assure the number of particles did not affect any of the model parameter tuning results.

To improve the computational efficiency of EczemaPF, the number of particles was decreased and its impact on the predictive performance of each severity sign was assessed. Full results are shown in Appendix B.2. It was concluded that all models had negligible changes when the number of particles was changed from 5000 to 1000. The framework was trained for all patients every day, for 80 days (duration of the dataset). For 5000 particles, this process took 6 hours. For 1000 particles, it took only about 1 hour and 30 minutes, corresponding to a 75% decrease in training time compared to the training with 5000 particles. Future work should investigate whether it is possible to further decrease the number of particles without worsening performance.

3.1.5 Parameter tuning limitations

While the parameter tuning results aforementioned provide an indication of the scale of the optimal parameters for EczemaPF, a more rigorous analysis should be undertaken. In particular, cross-validation should be performed as part of the tuning process. Furthermore, a different dataset could be used as a test dataset, so as to assess the performance of specific parameter combinations on a different data distribution.

In particular, the extent tuning results seem to be particularly unreliable, as they overfit to the data considered in this project. Visual analysis of the extent trajectories of different patients show large periods of constant extent, which thus explain why a more constant latent model (where the particles do not transition much each time-step) achieves better results. However, this choice of parameter is bound to fail for patients where extent dynamics are highly variable, regardless of the presence of possible flares. This can observed in the two example trajectories in Figure 9, where EczemaPF fails to predict larger changes in extent.

The optimal set of parameters found, which is used in the following results subsections, is reported in Appendix A.

3.2 Prediction of severity items

Figure 10 shows the learning curves of each of the particle filters of EczemaPF and the same curves for EczemaPred's training. EczemaPF managed to learn the dynamics of each severity

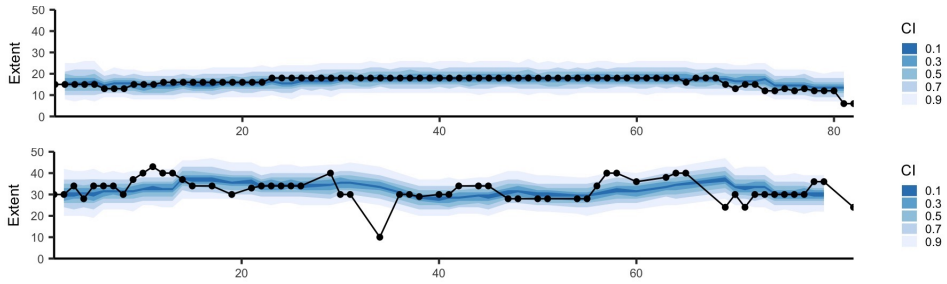


Figure 9: Extent prediction by tuned EczemaPF for two patients from the PFDC dataset. The confidence intervals represent the predictive distributions, and black dots represent the recordedd extent. The model was trained every 4 days, and new 4-day ahead predictions issued.

item as more data came in. However, both frameworks show predictions of extent and subjective signs appear to be more difficult than the other severity items, as shown by the much smaller lpd values. Whereas the former have lpds hovering around log(0.1), the latter reach much higher lpds, with some signs (thickening,dryness and swelling) achieving a mean lpd of log(0.9), showing a very high predictive performance.

When compared to EczemaPred, EczemaPF shows a significantly worse performance in the first training days, with its performance catching up as more data is given. This suggests EczemaPF seems to converge to EczemaPred's predictive performance as more data is given. However, a more through analysis, with different and larger datasets, is necessary to draw conclusions. It should also be analysed whether this is due to poorly chosen priors for EczemaPF, or whether there is an intrinsic aspect of state-estimation and parameter-estimation settings that explains this initial difference. For example, if we consider day 10, while each filter of EczemaPF (performing state-estimation) has only considered 10 datapoints (the data of a specific patient) to guess the latent severity sign, each model of EczemaPred (performing both state and parameter-estimation) has considered 10 (days) $*$ 16 (patients) datapoints, as it uses information from all patients to predict the severity items of each patient.

There are significant differences between the EczemaPred and EczemaPF's models which must be taken into account when comparing predictive performance between the two frameworks. These differences can be attributed to difficulties in the implementation of certain features with Particle filters. Firstly, EczemaPred not only learns the latent severity items, but aims to find the models parameters defining the dynamics of each severity item. Some of these parameters are defined for the entire population (set of patients), while others are patient-specific. Thus, EczemaPred's models are hierarchical and patient-specific. On the other hand, EczemaPF's

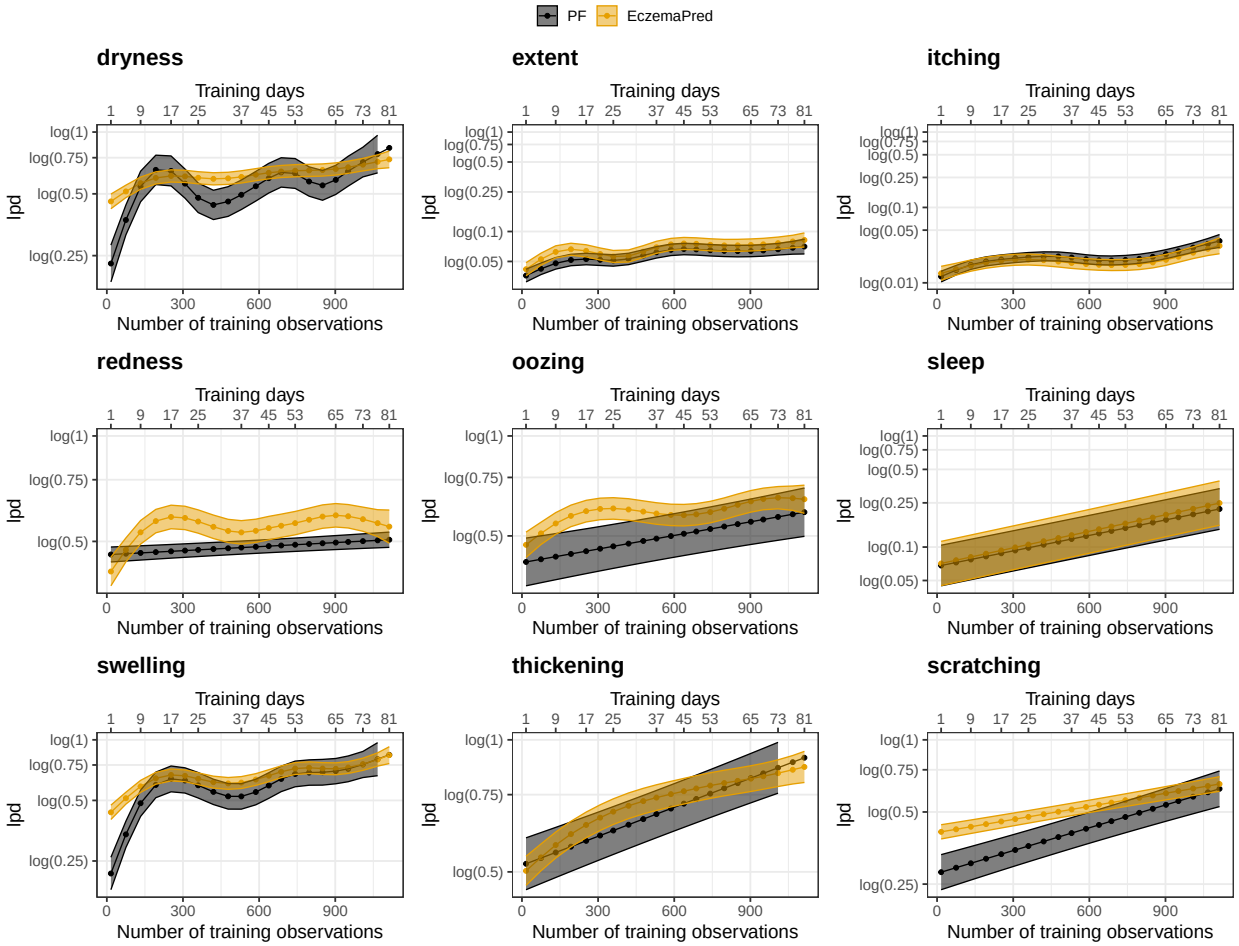


Figure 10: Learning curves for 4-days ahead prediction of both frameworks as a function of the number of training days/training observations.

Particle filters are all used in a state-estimation setting. That is, the parameters defining the latent dynamics and measurement distributions are fixed, and each filter only aims to track its latent severity item. Thus, while EczemaPF's predictions are patient-specific, the framework currently does not learn from other patients' data, or consider eczema dynamics that differ between patients.

The similarity of results towards the last training data points seems to suggest patient-specific models do not provide the predictive improvements that would have been expected from the very heterogeneous eczema severity data. However, the use of different distributions, as well as the difference between the state-estimation and parameter-estimation settings of the two frameworks, mean conclusions regarding the relevancy of patient-specific models cannot be taken.

Future work should focus on adapting EczemaPF's models and settings to mimic EczemaPred

fully, so results can be compared when the only difference is the inference algorithm used (PF vs HMC). Furthermore, EczemaPF should be compared to reference models, such as uniform forecast and historical forecasts, as well as to standard time-series forecasting models, such as AR models. This would allow for a better understanding of some of the possible drawbacks of Particle filters, and whether this algorithm could fully substitute HMC in eczema severity prediction.

3.3 PO-SCORAD prediction

To obtain PO-SCORAD predictions, the predictions for individual severity items (presented in the previous section) were aggregated each day according to the PO-SCORAD formula. Figures 11 and 12 each show the 4-day-ahead forward-chaining predictive trajectories, for individual severity signs as well as PO-SCORAD, of a patient using both EczemaPred and EczemaPF.

A visual analysis of the curves show most signs' predictions are almost identical for EczemaPred and EczemaPF. However, both extent and redness present starker differences that can be explained by the different distributions and parameters used for EczemaPF and described in section 3.1.

Firstly, the extent prediction by EczemaPF appears much more certain, with narrower confidence intervals. This is due to the choice of standard deviation for the extent latent dynamics (0.01, as described in 3.3.), which is quite small, resulting in a latent probability that can only very slightly change between time-points. This choice of parameter does obtain better lpd, as many patients showcase relatively flat extent dynamics, but fails to model any extent flares. However, this limitation is also common to EczemaPred, and thus is likely due to the models chosen, and not the parameters.

On the other hand, while both frameworks mostly agree in redness predictions, EczemaPF's confidence intervals are wider than EczemaPred's (shown by the lighter colours). Parameter tuning in section 3.1. shows that a larger standard deviation for both intensity model Gaussians results in better redness prediction, and thus this larger variance explains the wider confidence intervals observed in prediction curves. As the intensity observation model differs between the two frameworks, it is challenging to draw conclusions on the reasons behind the differences in curves.

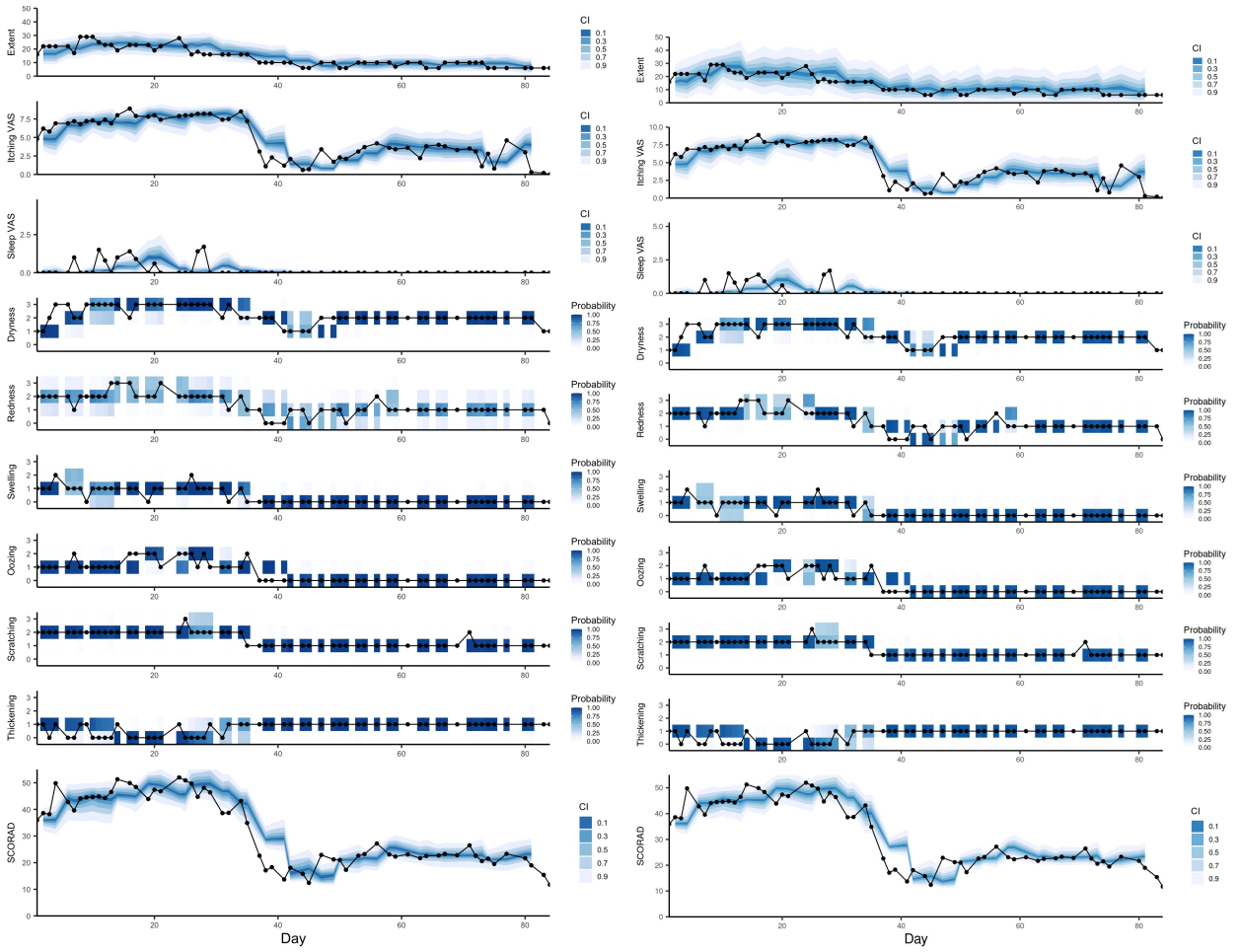


Figure 11: PO-SCORAD prediction (and breakdown for each severity item) by EczemaPF (left) and EczemaPred (right) for a patient in the PFDC dataset. Coloured intervals indicate stacked prediction intervals of higest density. Black dots represent the recorded severity item/PO-SCORAD. The modals were trained in a forward-chaining setting, and thus trained (and new 4-day ahead predictions issues) every 4 days.

3.4 Computational efficiency

Focusing on the main aim of this project, the computational efficiency of EczemaPF was studied, in particular when compared to EczemaPred's. To do so, both frameworks were trained and forecasted 4 days ahead for each day present in the dataset. The time taken to train the model and use it to forecast was analysed. As seen in Figure 13, all EczemaPred models became significantly slower as more datapoints were considered. On the contrary, EczemaPF's models run time did not change with the number of training observations. Furthermore, EczemaPF is faster for any number of datapoints for extent, itching and sleep, while it requires about 150 datapoints (or about 9 days of training for each patient) to become faster than EczemaPred for the remaining signs.

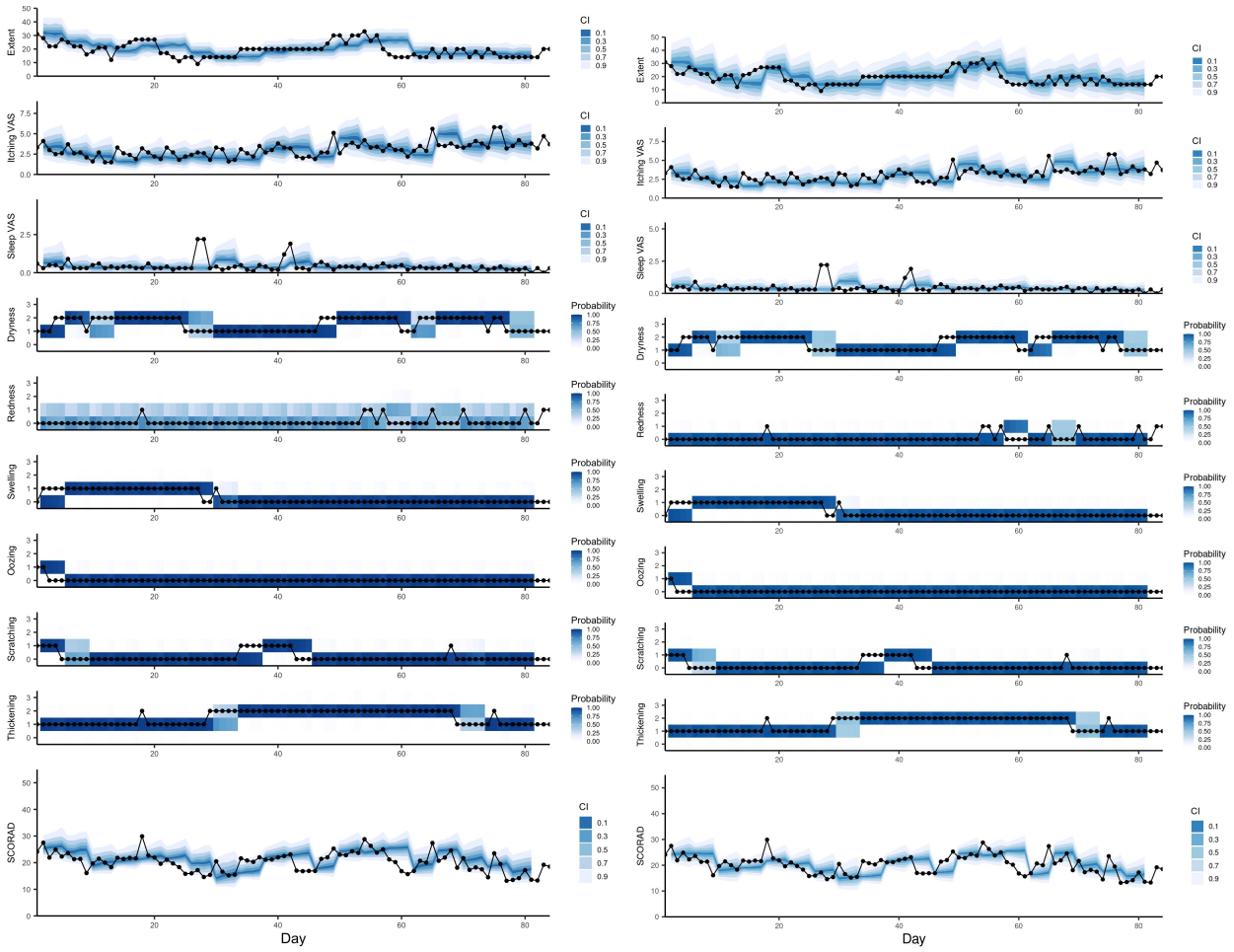


Figure 12: PO-SCORAD prediction (and breakdown for each severity item) by EczemaPF (left) and EczemaPred (right) for a patient in the PFDC dataset. Coloured intervals indicate stacked prediction intervals of higest density. Black dots represent the recorded severity item/PO-SCORAD. The modals were trained in a forward-chaining setting, and thus trained (and new 4-day ahead predictions issues) every 4 days.

In terms of real-world application of EczemaPF with its current models, each patient would wait on average 5 seconds for a PO-SCORAD prediction, or 1 minute and 15 seconds (for a 16-patient dataset) if other patients' data affects this prediction. The latter is already much faster than the average times for EczemaPred even with only 600 datapoints, as an analysis of Figure 5 shows EczemaPred would take on average about 10 minutes and 30 seconds for a prediction. This gap would only increase with the number of training datapoints.

It should be noted that while the run time of the particle filter each day is expected to stay independent of the number of training observations (characteristic of an online method), the implementation of a more complex model (to mirror EczemaPred), as well as the adaptation of the PFs to perform parameter estimation, are expected to significantly increase the run time. While it is unlikely the number of particles can be reduced further when these features

are implemented (if anything, parameter estimation is expected to require a larger number of particles), the PFs support a parallel implementation, where 2 of the 3 steps can be performed in parallel. Overall, EczemaPF is still expected to remain computationally more efficient, but the number of training observations at which it becomes more efficient is likely to increase. In a real-world application, where the number of training observations would grow boundlessly as days go on, the particle filter would still provide a more sustainable solution.

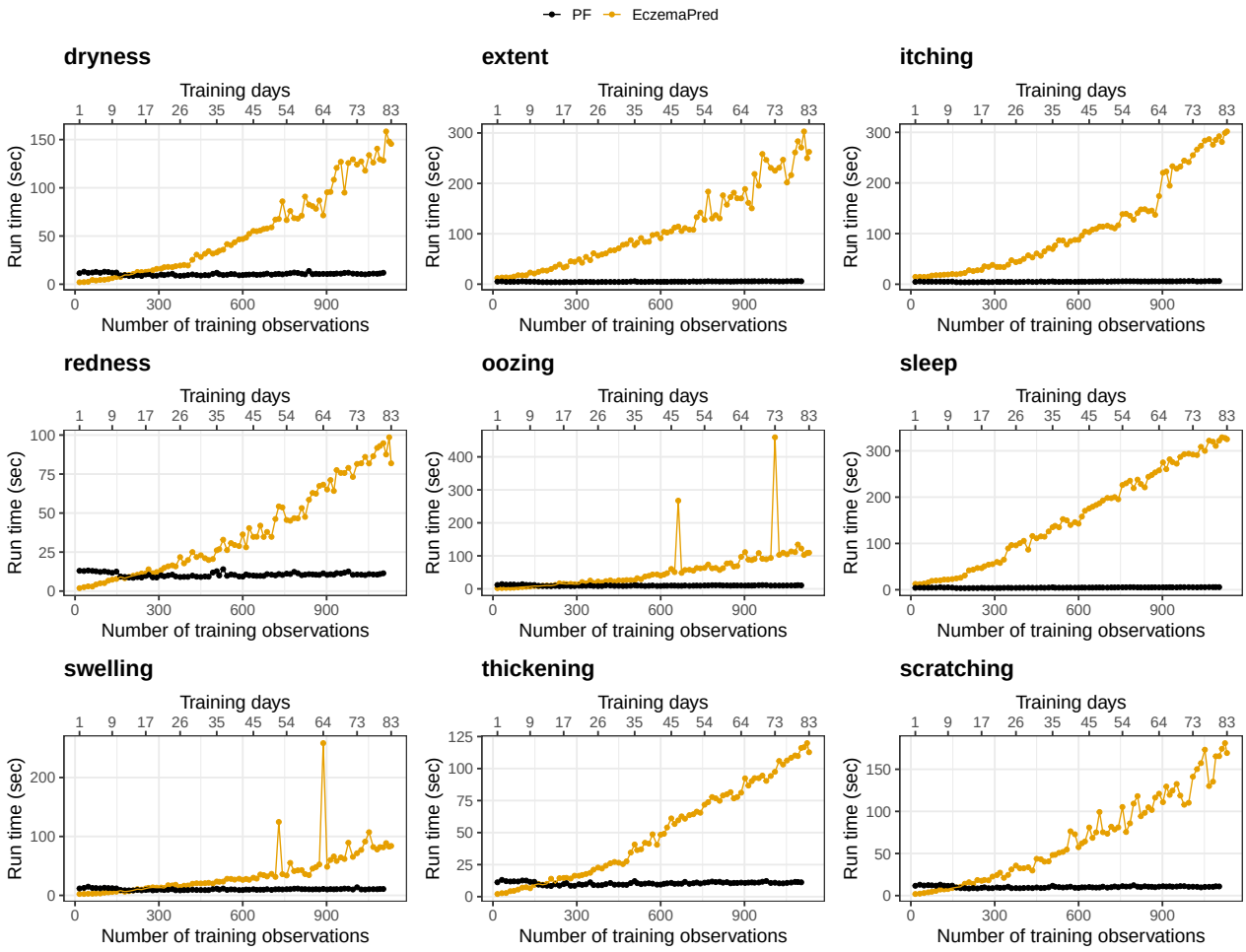


Figure 13: Run time as a function of the number of training observations for EczemaPF and EczemaPred. EczemaPF used the entire patients' pool data every day, as EczemaPred, to allow for a fair comparison.

3.5 Future steps

While EczemaPF computationally outperforms EczemaPred while achieving a comparable predictive performance for enough training datapoints, any comparison between the two frameworks is limited by the differences in the models. Thus, future work should focus on making

sure both frameworks perform exactly the same task (parameter and state estimation) using the same models of severity item dynamics. Furthermore, EczemaPF should be compared to reference models and standard time-series forecasting models. Only after this can the potential of PFs to replace MCMC be fully assessed.

Notwithstanding, EczemaPF's predictive performance does suggest perhaps EczemaPred's model added complexity (in the form of more complex distributions and a hierarchical model outputting patient-specific predictions) brings reduced benefits. Thus, an alternative approach, where EczemaPred's models are simplified to match EczemaPF could also allow for a better comparison of the two methods, as well as a better understanding of the models required to accurately describe eczema severity dynamics. However, this analysis would once again require different datasets, so as to understand whether patient-specific models' benefits are clearer for different data distributions.

Furthermore, the comparison between frameworks was simplified in this project. Future work should aim to compare these frameworks across different datasets and different forecasting times (how many days ahead the frameworks are predicting).

4 Conclusion

In summary, in this report I introduced EczemaPF, a computational framework aimed at predicting eczema severity dynamics. This framework, unlike its predecessor EczemaPred, allows for near real-time prediction, with patients waiting around 5 seconds (per patient in the dataset) for a prediction. EczemaPF proved to match EczemaPred's performance as more data was available. Future work should focus on implementing EczemaPF with exactly the same models as EczemaPred, to better assess whether Particle filters could fully substitute MCMC in eczema severity prediction. Furthermore, EczemaPF should be adapted to perform parameter estimation to account for patient-specific eczema severity dynamics. In the future, EczemaPF could be used in conjunction with EczemaNet [9] to create a smartphone app used as a treatment tool, providing patients with a likely evolution of their symptoms as well as appropriate treatment suggestions.

References

[1] Stephan Weidinger, Lisa A. Beck, Thomas Bieber, Kenji Kabashima, and Alan D. Irvine. Atopic dermatitis. *Nature Reviews Disease Primers*, 4:1–20, 06 2018. doi: 10.1038/s41572-018-0001-z.

[2] Thomas Bieber, Angelo M. D'Erme, Cezmi A. Akdis, Claudia Traidl-Hoffmann, Roger Lauener, Georg Schäppi, and Peter Schmid-Grendelmeier. Clinical phenotypes and endophenotypes of atopic dermatitis: Where are we, and where should we go? *Journal of Allergy and Clinical Immunology*, 139:S58–S64, 04 2017. doi: 10.1016/j.jaci.2017.01.008.

[3] S Tofte, M Graeber, R Cherill, M Omoto, M Thurston, and JM. Hanifin. Eczema area and severity index (easi): A new tool to evaluate atopic dermatitis. *Journal of the European Academy of Dermatology and Venereology*, 11:S197, 09 1998. doi: 10.1016/s0926-9959(98)95291-6.

[4] Jochen Schmitt, Phyllis I. Spuls, Kim S. Thomas, Eric Simpson, Masutaka Furue, Stefanie Deckert, Magdalene Dohil, Christian Apfelbacher, Jasvinder A. Singh, Joanne Chalmers, and Hywel C. Williams. The harmonising outcome measures for eczema (home) statement to assess clinical signs of atopic eczema in trials. *Journal of Allergy and Clinical Immunology*, 134:800–807, 10 2014. doi: 10.1016/j.jaci.2014.07.043.

[5] JF Stalder, A Taieb, DJ Atherton, and al et. Severity scoring of atopic dermatitis: The scorad index. *Dermatology*, 186:23–31, 1993. doi: 10.1159/000247298.

[6] Guillem Hurault, Jean François Stalder, Sophie Mery, Alain Delarue, Markéta Saint Aroman, Gwendal Josse, and Reiko J. Tanaka. Eczemapred: A computational framework for personalised prediction of eczema severity dynamics. *Clinical and Translational Allergy*, 12, 03 2022. doi: 10.1002/clt2.12140.

[7] P. P. Vakharia, R. Chopra, R. Sacotte, N. Patel, S. Immaneni, T. White, R. Kantor, D. Y. Hsu, and J. I. Silverberg. Validation of patient-reported global severity of atopic dermatitis in adults. *Allergy*, 73:451–458, 02 2018. doi: 10.1111/all.13309.

[8] Yael A. Leshem, Joanne R. Chalmers, Christian Apfelbacher, Masutaka Furue, Louise A.A. Gerbens, Cecilia A.C. Prinsen, Jochen Schmitt, Phyllis I. Spuls, Kim S. Thomas, Hywel C. Williams, and Eric L. Simpson. Measuring atopic eczema symptoms in clinical practice:

The first consensus statement from the harmonising outcome measures for eczema in clinical practice initiative. *Journal of the American Academy of Dermatology*, 82:1181–1186, 05 2020. doi: 10.1016/j.jaad.2019.12.055.

[9] Kevin Pan, Guillem Hurault, Kai Arulkumaran, Hywel C. Williams, and Reiko J. Tanaka. Eczemanet: Automating detection and severity assessment of atopic dermatitis. *Machine Learning in Medical Imaging*, pages 220–230, 2020.

[10] Pierre Fabre Dermo Cosmetique. Collection of clinical and biometrological data in adult subjects suffering from atopic dermatitis (ad) during a three-month period, 09 2020.

[11] Andrew Gelman, Jessica Hwang, and Aki Vehtari. Understanding predictive information criteria for bayesian models. *Statistics and Computing*, 24:997–1016, 08 2013. doi: 10.1007/s11222-013-9416-2.

[12] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings F - Radar and Signal Processing*, 140:107–113, 04 1993. doi: 10.1049/ip-f-2.1993.0015.

[13] Simone Zocca, Yihan Guo, Alex Minetto, and Fabio Dovis. Improved weighting in particle filters applied to precise state estimation in gnss. *Frontiers in Robotics and AI*, 9, 08 2022. doi: 10.3389/frobt.2022.950427.

[14] S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso, and A. M. Stuart. Importance sampling: Intrinsic dimension and computational cost. *Statistical Science*, 32, 08 2017. doi: 10.1214/17-sts611.

[15] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50:174–188, 2002. doi: 10.1109/78.978374.

[16] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics*. MIT Press, 08 2005.

[17] Fredrik Gustafsson. Particle filter theory and practice with positioning applications. *IEEE Aerospace and Electronic Systems Magazine*, 25:53–82, 07 2010. doi: 10.1109/maes.2010.5546308.

[18] Jos Elfring, Elena Torta, and René van de Molengraft. Particle filters: A hands-on tutorial. *Sensors*, 21:438, 01 2021. doi: 10.3390/s21020438.

[19] Nikolas Kantas, Arnaud Doucet, Sumeetpal S. Singh, Jan Maciejowski, and Nicolas Chopin. On particle methods for parameter estimation in state-space models. *Statistical Science*, 30:328–351, 08 2015. doi: 10.1214/14-sts511.

[20] Daniel M. Sheinson, Jarad Niemi, and Wendy Meiring. Comparison of the performance of particle filter algorithms applied to tracking of a disease epidemic. *Mathematical Biosciences*, 255:21–32, 09 2014. doi: 10.1016/j.mbs.2014.06.018.

[21] Matthew Hoffman, David Blei, Chong Wang, John Paisley, Jpaisley@berkeley Edu, and Tommi Jaakkola. Stochastic variational inference. *Journal of Machine Learning Research*, 14:1303–1347, 2013.

[22] Corey Montella. The kalman filter and related algorithms: A literature review, 05 2011.

[23] Nick Foti, Jason Xu, Dillon Laird, and Emily Fox. Stochastic variational inference for hidden markov models, 2014.

[24] Robert Moss, David J. Price, Nick Golding, Peter Dawson, Jodie McVernon, Rob J. Hyndman, Freya M. Shearer, and James M. McCaw. Forecasting covid-19 activity in australia to support pandemic response: May to october 2020. *Scientific Reports*, 13:8763, 05 2023. doi: 10.1038/s41598-023-35668-6.

[25] Robert Moss, Alexander Zarebski, Sandra Carlson, and James McCaw. Accounting for healthcare-seeking behaviours and testing practices in real-time influenza forecasts. *Tropical Medicine and Infectious Disease*, 4:12, 01 2019. doi: 10.3390/tropicalmed4010012.

A EczemaPF models

In this appendix, the following notation will be followed:

- $y(t)$ is the observed severity sign at time t.
- $\mathcal{N}(\mu,\sigma^2)$ is a Gaussian distribution with mean μ and variance σ^2.

A.1 Extent models

As in EczemaPred [6], the extent model assumes that we can divide a patient's skin into 100 patches, each with a probability p of being lesional. This latent probability p is modelled by a Gaussian random walk, whose standard deviation was tuned as described in section 3.3.1 and determined to be 0.01. The measurement is defined as a binomial distribution with 100 trials, corresponding to a patient counting the number of lesional patches in their skin.

$$\textbf{Latent model}\text{: } p(t+1) \sim \mathcal{N}(p(t), 0.001)$$

$$\textbf{Observation model}\text{: } y(t) \sim \mathcal{B}(100, p(t))$$

with the prior:

$$p \sim \mathcal{N}(0.5, 0.25)$$

The prior was chosen to be weakly informative. Further study should consider centering this value on lower values of probability, as observation of extent data shows it only very rarely goes beyond values of 40. This model is replicated for all patients, but the particle filter tracks a (possibly) different latent probability for each patient.

A.2 Intensity models

Unlike for the two other types of severity signs, in this case the latent variable does not represent a probability. For the intensity signs, both the latent and measurement dynamics were modelled as Gaussian random walks, representing the dynamics of the latent intensity sign

and its observation respectively. The standard deviations of these Gaussians were tuned, as described in 3.3.2, for each intensity sign. Furthermore, the prior on the latent severity was also tuned for each intensity sign. As each intensity sign takes a discrete ordinal value from 0 to 3, each observation y was rounded to its closest integer.

The priors were centered around 1 as it was observed the intensity signs rarely take the value of 3. The models follow the general formula below.

$$\textbf{Latent model}: p(t+1) \sim \mathcal{N}(p(t), {\sigma_1}^2)$$

$$\textbf{Observation model}: y(t) \sim \mathcal{N}(p(t), {\sigma_2}^2)$$

with the prior:

$$p \sim \mathcal{N}(1, {\sigma_3}^2)$$

The standard deviations of each model (shown in Table 5) were determined based on the results on Appendix B.

Table 5: Tuned parameters and priors for different intensity signs.

Sign	**Latent model** σ_1	**Observation model** σ_2	**Prior**
Dryness	0.1	0.1	$p \sim \mathcal{N}(1, 1)$
Redness	0.2	0.3	$p \sim \mathcal{N}(1, 0.25)$
Swelling	0.1	0.1	$p \sim \mathcal{N}(1, 0.25)$
Oozing	0.1	0.1	$p \sim \mathcal{N}(1, 1)$
Thickening	0.1	0.1	$p \sim \mathcal{N}(1, 0.25)$
Scratching	0.1	0.1	$p \sim \mathcal{N}(1, 0.25)$

A.3 Subjective models

For the subjective signs, a latent probability followed a Gaussian random walk, with its observation being described by a binomial distribution with 100 trials, as is also the case with extent. However, unlike the latter, for the subjective signs this random walk was performed

on a logit scale, with a standard deviation resulting from a tuning process described in 3.3.3. Furthermore, the observation (the number of successful Bernoulli trials) is divided by 10, so as to provide a discrete subjective sign value from 0 to 10, with resolution of 0.1.

$$\textbf{Latent model}\text{: } logit(p(t+1)) \sim \mathcal{N}(logit(p(t)), 0.09)$$

$$\textbf{Observation model}\text{: } y(t) \sim \mathcal{B}(100, p(t))$$

with the prior:

$$p \sim \mathcal{N}(0.5, 0.25)$$

Alike the prior for the extent, this prior is weakly informative, and further study should explore a weakly informative prior centered at lower values of probability. This model is replicated for all patients, with the particle filter tracking a different latent probability distribution for each patient.

B Supplementary results

B.1 Intensity signs

Four different trials were performed, with different standard deviations and priors on the latent variable, and the performance of the particle filter associated with each intensity sign was assessed. The experimental conditions in each trial are shown in Table 6, and the results of these trials are shown in Table 7.

Table 6: Experimental conditions for each trial of the intensity signs' tuning process.

Trial	**Latent model** σ_1	**Observation model** σ_2	**Prior**
1	0.1	0.1	$p \sim \mathcal{N}(1, 0.25)$
2	0.05	0.1	$p \sim \mathcal{N}(1, 0.25)$
3	0.2	0.3	$p \sim \mathcal{N}(1, 0.25)$
4	0.1	0.1	$p \sim \mathcal{N}(1, 1)$

Table 7: Lpd for each trial of the intensity signs' tuning process, for different intensity signs.

Sign	**Trial 1 lpd**	**Trial 2 lpd**	**Trial 3 lpd**	**Trial 4 lpd**
Dryness	log(0.73)	log(0.51)	log(0.61)	log(0.61)
Redness	log(0.49)	log(0.35)	log(0.51)	log(0.5)
Swelling	log(0.73)	log(0.64)	log(0.61)	log(0.73)
Oozing	log(0.57)	log(0.41)	log(0.56)	log(0.61)
Thickening	log(0.85)	log(0.76)	log(0.65)	log(0.84)
Scratching	log(0.53)	log(0.39)	log(0.55)	log(0.54)

The observation process is assumed to be more uncertain than the latent severity dynamics, and thus in both trials where the standard deviations were changed, the observation model standard deviation was chosen to be higher. The difference between increasing the latent model standard deviation, as opposed to the observation model's, was not investigated.

When values of lpd at 65 days were very close, the learning curves were analysed to choose the

ones with best performance at 80 days (end of trials), as well as average performance throughout the trial.

Based on the results of Table 7, and an analysis of the learning curves, the chosen parameters and priors for each intensity sign were as shown in Table 5 (Appendix A).

B.2 Number of particles

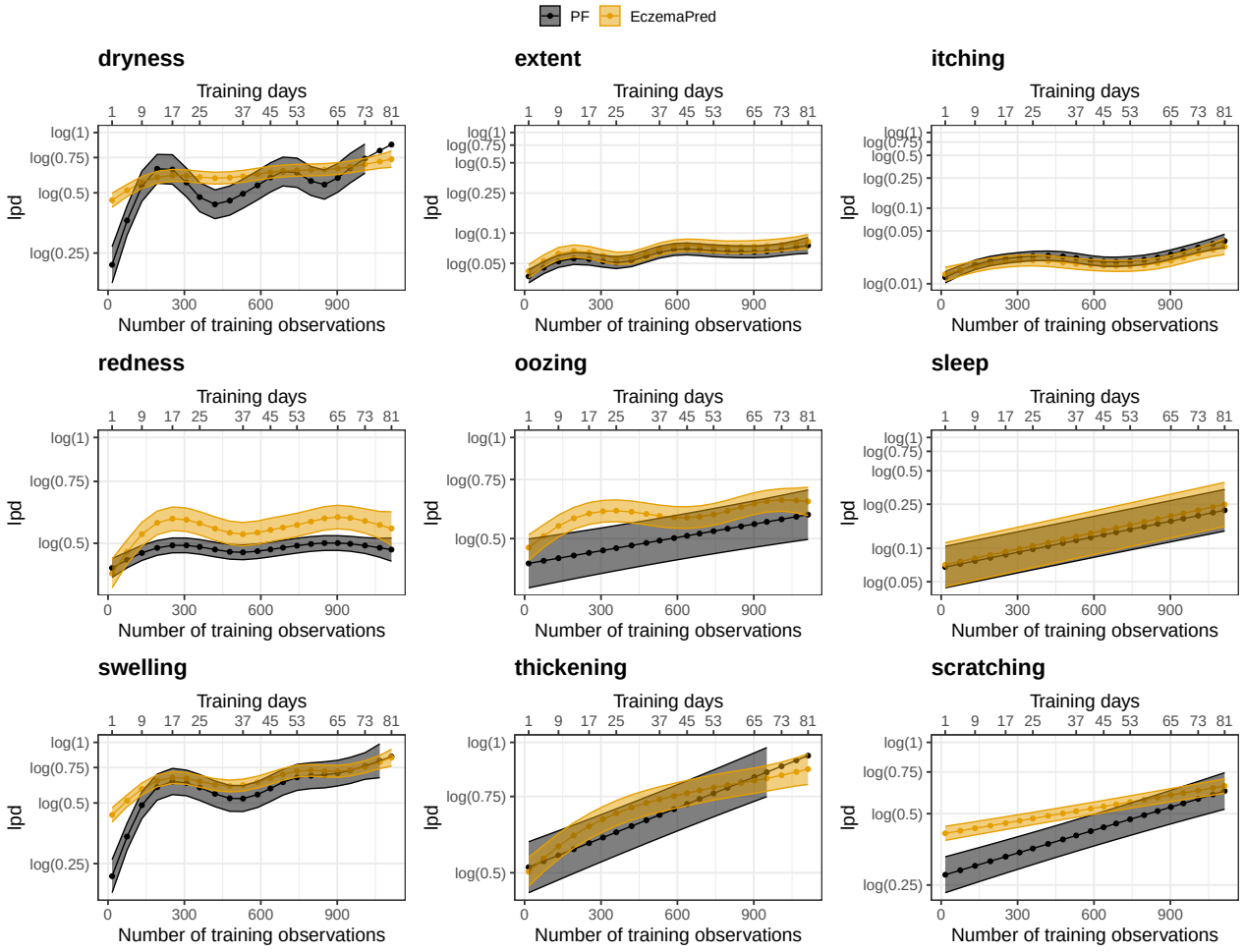


Figure 14: Learning curves for 4-days ahead prediction of both frameworks as a function of the number of training days/training observations, using 5000 particles.

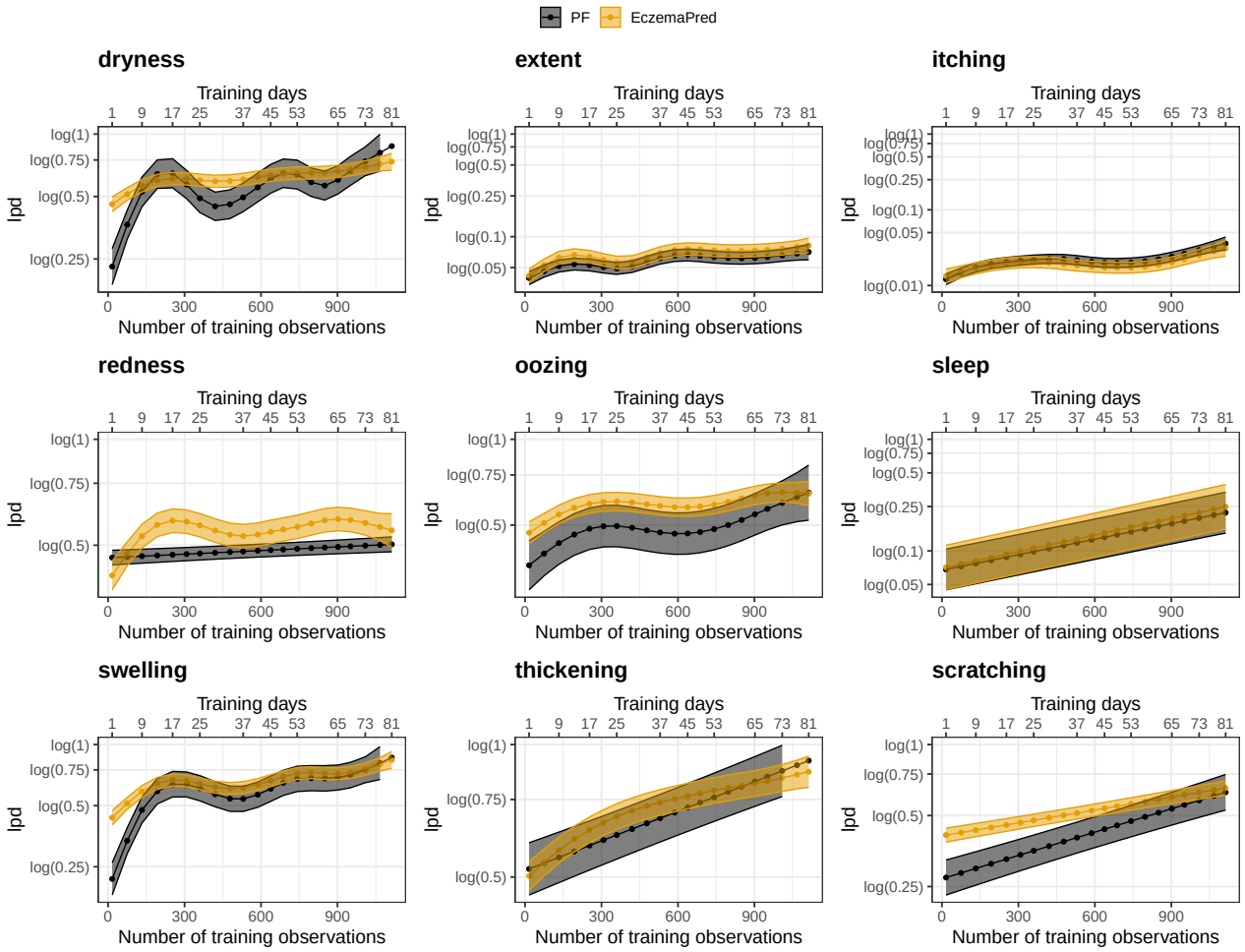


Figure 15: Learning curves for 4-days ahead prediction of both frameworks as a function of the number of training days/training observations, using 1000 particles.