

1 A Revised Implicit Equal-Weights Particle Filter

2 Jacob Skauvold¹, Jo Eidsvik¹, Peter Jan van Leeuwen^{2,3}, and
3 Javier Amezcua^{2,3}

4 ¹Norwegian University of Science and Technology

5 ²Data Assimilation Research Centre, University of Reading, UK

6 ³National Centre for Earth Observation, UK

7 January 14, 2019

8 **Abstract**

9 Particle filters are fully non-linear data assimilation methods and as
10 such are highly relevant. While the standard particle filter degenerates
11 for high-dimensional systems, recent developments have opened the way
12 for new particle filters that can be used in such systems.

13 The implicit equal-weights particle filter (IEWPF) is an efficient ap-
14 proach which avoids filter degeneracy because it gives equal particle weights
15 by construction. The method uses implicit sampling whereby auxiliary
16 vectors drawn from a proposal distribution undergo a transformation be-
17 fore they are added to each particle.

18 In the original formulation of the IEWPF, the proposal distribution
19 has a gap causing all but one particle to have an inaccessible region in
20 state space. We show that this leads to a systematic bias in the pre-
21 dictions and we modify the proposal distribution to eliminate the gap.
22 We achieved this by using a two-stage proposal method, where a single

23 variance parameter is tuned to obtain adequate statistical coverage prop-
24 erties of the predictive distribution. We discuss properties of the implicit
25 mapping from an auxiliary random vector to the state vector, keeping in
26 mind the aim of avoiding particle resampling. The revised filter is tested
27 on linear and weakly nonlinear dynamical models in low-dimensional and
28 moderately high-dimensional settings, demonstrating the success of the
29 new methodology in removing the bias.

30 **1 Introduction**

31 Geophysical models involving numerical simulations of processes unfolding in
32 space and time often take the form of state space models with non-linear dy-
33 namics and millions of state variables. As the evolution of such systems is
34 sensitive to initial conditions and boundary conditions, which are almost never
35 known precisely, the actual system state is generally uncertain. Model error,
36 failure of the numerical model to faithfully represent the simulated process, also
37 contributes to system state uncertainty. If observations of the modelled sys-
38 tem are available, then incorporating information from these into the model
39 through data assimilation can mitigate uncertainty and lead to more accurate
40 predictions.

41 Data assimilation in a Bayesian setting begins with a prior probability distri-
42 bution representing background knowledge about the unknown state variables.
43 The relationships between states and observations are represented by condi-
44 tional probability distributions referred to as the likelihood. Combining the
45 prior distribution and likelihood according to Bayes' theorem yields a posterior
46 distribution of the state conditional on the observations. When this is done over
47 time, data assimilation conditions the dynamical model to data.

48 Variational data assimilation methods like 3D-Var and 4D-Var use optimi-

49 sation to locate the posterior mode (Asch et al., 2016; Van Leeuwen et al., 2015;
50 Fletcher, 2017). While variational data assimilation methods do not necessarily
51 characterise the spread of the posterior distribution, an estimate of the poste-
52 rior covariance is available via the inverse of the Hessian evaluated at the mode.
53 More direct uncertainty quantification is possible with ensemble-based data as-
54 simulation methods such as the many variants of the ensemble Kalman filter
55 (EnKF), see e.g. Evensen (2009). However, the EnKF uses linear updating and
56 implicitly assumes that the state distribution and likelihood are Gaussian. This
57 limits the applicability of EnKF variants to only mildly non-linear dynamical
58 models.

59 Particle filters (PFs), see e.g. Doucet et al. (2000), most of which are based
60 on importance sampling, have no assumptions of linearity or Gaussianity. They
61 work by propagating particles, or model realizations, forward in time via a
62 forecast step and then weighting particles according to the likelihood, so that
63 the resulting weighted ensemble of particles represents the posterior probability
64 density. Some PF variants modify the forecast step by drawing particles from
65 a proposal distribution instead of the forward model (e.g. Doucet et al., 2000;
66 Van Leeuwen, 2009; Morzfeld et al., 2012; Van Leeuwen et al., 2015). This is then
67 accounted for in the weighting step. PFs are appealing in large part because
68 they are free of distributional assumptions and will, given enough particles,
69 correctly sample the posterior distribution even when applied to highly non-
70 linear dynamical models. In practice, when the number of particles is limited,
71 PFs are subject to the curse of dimensionality and can be relied on for correct
72 sampling only when state and observation dimensions are low to moderate.
73 Applied to high-dimensional data assimilation tasks, PFs tend to suffer from
74 filter degeneracy in the form of sample impoverishment. That is, the distribution
75 of particle weights, which is initially uniform, quickly begins to concentrate

76 around, and eventually collapses onto, a small subset of particles, effectively
77 reducing the ensemble size (Snyder et al., 2008).

78 There have been several approaches trying to combine strengths from PFs
79 with EnKF approaches. Stordal et al. (2011) constructed a useful Gaussian
80 mixture approximation to the predicted distribution at each step, bridging the
81 EnKF update with a special kind of PF updates. Rezaie and Eidsvik (2012)
82 shrunked the PF update towards the EnKF update, also relying on Gaussian
83 mixture models, and tuned the shrinkage parameter to avoid degeneracy while
84 maintaining reasonable statistical properties. Frei and Künsch (2013) applied
85 a tuning parameter in the exponent of the likelihood part, where parts of the
86 data (with larger variance) are used in an EnKF update, while the remaining
87 part is used in a PF step. In principle, these approaches have the non-linear
88 appeal of PFs, but automatized tuning tends to give results closer to the EnKF
89 output for high dimensional systems and moderate particle sizes (Stordal et al.,
90 2011).

91 Although in theory filter degeneracy issues can be remedied by increasing the
92 number of particles, computational limitations restrict ensemble sizes to around
93 100 particles in many data assimilation applications (Van Leeuwen, 2009). What
94 is desired in such cases is a PF variant that is resistant to filter degeneracy and
95 maintains a resonable particle weight distribution when applied to nonlinear
96 dynamical systems. Unlike the standard PF, such a filter might be a viable
97 solution for nonlinear and high dimensional data assimilation despite having to
98 operate with only a moderate number of particles.

99 The equivalent weights particle filter (EWPF, Van Leeuwen, 2010; Ades and
100 Van Leeuwen, 2013) is a non-linear data assimilation approach which uses a
101 proposal distribution constructed to give equal weights in the update step, thus
102 avoiding particle degeneracy. Depending on the specifics of the proposal density

103 used, some or all particles may need to be resampled to maintain exact equality
104 between weights.

105 The implicit equal-weights particle filter (IEWPF), introduced by Zhu et al.
106 (2016), similarly prevents filter degeneracy by constructing the proposal dis-
107 tribution so that the weights are uniform. The IEWPF combines the implicit
108 sampling framework of Chorin et al. (2013) with the equal-weights idea from
109 Ades and Van Leeuwen (2013). By the implicit construction no parameter tun-
110 ing is required. However, the approach tends to give biased results, particularly
111 for moderate state dimensions, because its construction yields a proposal density
112 for particle updates that is zero on parts of state space.

113 The new contributions of the current paper are first a demonstration that
114 this bias is systematic, and leads to underestimation of the filter variance. Sec-
115 ondly, we modify the IEWPF to remedy some of the deficiencies of the proposal
116 distribution under the original IEWPF formulation, specifically to eliminate the
117 gap in state space described by Zhu et al. (2016) and to reduce the mismatch
118 between the reported and actual prediction variance of the ensemble represen-
119 tation of the posterior probability density. Our suggested modification achieves
120 this by introducing an additional perturbation of each particle in the update
121 step of the filtering algorithm. Adjusting the scale of this perturbation enables
122 calibration of ensemble spread without compromising particle weight equality.
123 Additionally, the revised IEWPF can be applied to systems of any dimension.
124 This is in contrast to the original IEWPF, which relied on an approximation
125 that is only valid when the state dimension is large.

126 The new filter is a substantial improvement of the original IEWPF as it
127 provides a way to mitigate the bias in the original method. Still, it should
128 be noted that as with the original IEWPF, the emphasis is on handling non-
129 Gaussianity resulting from a nonlinear dynamical model rather than a non-

130 Gaussian likelihood. Allowing a nonlinear observation operator does not pose
 131 a fundamental problem but part of the analytical development involving the
 132 incomplete γ functions would not be possible, and solutions to the nonlinear
 133 equations would rely more on iterative methods. Such an extension is outside
 134 the scope of the this work.

135 This paper is organized as follows: In Section 2 the original single-stage
 136 IEWPF algorithm is described. In Section 3 the new two-stage IEWPF is
 137 presented. In Section 4 a linear example and a non-linear Lorenz96 example
 138 are studied.

139 2 Implicit equal-weights particle filter (IEWPF)

140 In this section we describe the main ideas and building blocks of the IEWPF
 141 algorithm. Some properties and challenges of this algorithm are discussed. A
 142 modified version of the filter is then described in Section 3.

143 2.1 Problem description and background

144 Consider a dynamical system with an N_x -dimensional state vector \mathbf{x}^n , $n =$
 145 $0, 1, \dots, n_t$. Set initial distribution $\mathbf{x}^0 \sim N(\boldsymbol{\mu}, \mathbf{B})$, denoting an N_x -dimensional
 146 Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix \mathbf{B} . Given the
 147 state at time t_{n-1} the state at time t_n is given by

$$\mathbf{x}^n = \mathcal{M}(\mathbf{x}^{n-1}) + \mathbf{u}^n, \quad (1)$$

148 where \mathcal{M} denotes forward integration of the dynamical system, and $\mathbf{u}^n \sim$
 149 $N(\mathbf{0}, \mathbf{Q})$ represents additive model error that we assume to be independent
 150 over time.

151 Suppose that at times $m \in \{1, 2, \dots\}$ an observation vector $\mathbf{y}^m \in \mathbb{R}^{N_y}$ is

152 available. The relationship between the state and observation vectors is

$$\mathbf{y}^m = \mathbf{H}\mathbf{x}^m + \mathbf{v}^m. \quad (2)$$

153 Here, \mathbf{H} is a size $N_y \times N_x$ linear observation operator and $\mathbf{v}^m \sim N(\mathbf{0}, \mathbf{R})$
154 represents additive observation error. In this article we will consider observation
155 operators which simply select certain elements of the state vector, but operations
156 like averaging and convolution of state vector elements are also possible. We
157 assume that the error terms are independent over time, and independent of the
158 error terms in the dynamical system model. Furthermore we will assume in the
159 remainder of this article that observations are available at every time step n , so
160 that the above notation may be simplified by letting $m = n$.

161 The filtering problem consists of estimating the current state \mathbf{x}^n given all
162 available observations up to time n . We denote the set of observations by $\mathbf{y}^{1:n}$.
163 The filtering probability density function is $p(\mathbf{x}^n | \mathbf{y}^{1:n})$, and this is computed
164 sequentially for $n = 1, 2, \dots$. The PF (Gordon et al., 1993) represents the
165 filtering distribution at every stage n by a size N_e ensemble of state realizations
166 \mathbf{x}_i^n , $i = 1 \dots, N_e$, called particles. Weights w_i^n , $i = 1, \dots, N_e$, are assigned to
167 each particle. A particle's weight is proportional to the likelihood of all data
168 along its sample path. It is updated sequentially using the multiplicative factor
169 $p(\mathbf{y}^n | \mathbf{x}_i^n)$.

170 One major problem affecting PF methods is sample degeneracy, also known
171 as sample impoverishment. This happens when the distribution of weight over
172 particles becomes more unequal with every iteration. Eventually this leads to
173 a situation where almost all weight is concentrated on a single particle, so that
174 the effective sample size is much smaller than the nominal ensemble size, and
175 the usefulness of the resulting ensemble is very limited. To avoid this behaviour
176 it is of interest to minimise the variance of the weights with respect to the

177 filtering distribution. The weights could be reset by including various kinds
 178 of resampling of particles at different stages, but this is usually not enough to
 179 avoid degeneracy in high-dimensional state space models.

180 Some PF variants employ importance sampling (Van Leeuwen, 2009), whereby
 181 particle updates are drawn from a proposal probability density function, or im-
 182 portance function, $q(\mathbf{x})$. Proposal densities are typically chosen to allow easy
 183 sampling and pointwise evaluation. The choice of proposal distribution can also
 184 affect the overall efficiency of the algorithm. For consistent results the particle
 185 weights are multiplied by the ratio of the target density to the proposal density

$$\frac{p(\mathbf{x}^n | \mathbf{y}^{1:n})}{q(\mathbf{x}^n)}. \quad (3)$$

186 According to Doucet et al. (2000), among potential importance functions of
 187 the form $q(\mathbf{x}^n) = q(\mathbf{x}^n | \mathbf{x}_i^{n-1}, \mathbf{y}^n)$, the one which minimises the particle weight
 188 variance is the conditional distribution $p(\mathbf{x}^n | \mathbf{x}_i^{n-1}, \mathbf{y}^n)$, referred to as the op-
 189 timal proposal density (OPD, Snyder et al., 2015). Ades and Van Leeuwen
 190 (2013) showed that a PF using the optimal proposal density as an importance
 191 function will degenerate slower than the standard SIR PF, but the exponential
 192 dependence on the size of the system remains the same.

193 With our modeling assumptions, this OPD is Gaussian. At stage n , and for
 194 every particle $i = 1 \dots, N_e$, its mean and covariance matrix, denoted by $\mathbf{x}_i^{n,a}$
 195 and \mathbf{P} respectively, are given by

$$\mathbf{x}_i^{n,a} = \mathcal{M}(\mathbf{x}_i^{n-1}) + \mathbf{QH}^T(\mathbf{HQH}^T + \mathbf{R})^{-1}(\mathbf{y}^n - \mathbf{HM}(\mathbf{x}_i^{n-1})), \quad (4)$$

196 and

$$\mathbf{P} = (\mathbf{Q}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1}. \quad (5)$$

197 **2.2 Single-stage IEWPF**

198 The proposal density can be constructed in various ways. We will now discuss
 199 the implicit scheme used in the IEWPF. This implicit sampling is realised by
 200 centering the proposal distribution on the mode of the OPD for each particle,
 201 and adding a random perturbation vector which is pre-multiplied by the square
 202 root of the OPD covariance matrix (5) and by a particle-specific scale factor
 203 $\alpha_i^{1/2}$.

204 Mathematically, the updated state of particle i is computed according to

$$\mathbf{x}_i^n = \mathbf{x}_i^{n,a} + \alpha_i^{1/2} \mathbf{P}^{1/2} \boldsymbol{\xi}_i^n, \quad (6)$$

205 where the random vector $\boldsymbol{\xi}_i^n \in \mathbb{R}^{N_x}$ is drawn from the proposal distribution
 206 $q(\boldsymbol{\xi}_i^n)$, which is specified as $N(\mathbf{0}, \mathbf{I}_{N_x})$. With $\alpha_i = 1$ this scheme is equivalent
 207 to drawing samples from the OPD. When $\alpha_i \neq 1$, the corresponding sampling
 208 distribution is either compressed or extended relative to the OPD. Note that α_i
 209 will change over time steps, but for notational convenience we have suppressed
 210 the superscript n .

By selecting α_i judiciously one can gain flexibility in the algorithm and avoid
 particle degeneracy, for instance by aiming for equal weights like we do here.
 The weight of particle i is given by

$$\begin{aligned} w_i^n &= w_i^{n-1} \frac{p(\mathbf{x}^n | \mathbf{x}_i^{n-1}) p(\mathbf{y}^n | \mathbf{x}^n)}{q(\boldsymbol{\xi}_i^n)} \left\| \frac{\partial \mathbf{x}^n}{\partial \boldsymbol{\xi}_i^n} \right\| \\ &= \frac{1}{N_e} \frac{p(\mathbf{x}^n | \mathbf{x}_i^{n-1}, \mathbf{y}^n) p(\mathbf{y}^n | \mathbf{x}_i^{n-1})}{q(\boldsymbol{\xi}_i^n)} \left\| \frac{\partial \mathbf{x}^n}{\partial \boldsymbol{\xi}_i^n} \right\|, \end{aligned} \quad (7)$$

211 where it is assumed that $w_i^{n-1} = 1/N_e$ for all particles i . To have equal weights
 212 $w_1^n = w_2^n = \dots = w_{N_e}^n = w_{\text{target}}^n$ the unnormalized log-weights must also be

213 equal, hence for each particle i the scalar α_i must satisfy

$$(\alpha_i - 1)\boldsymbol{\xi}_i^T \boldsymbol{\xi}_i - 2 \log(\alpha_i^{N_x/2}) - 2 \log \left(\left| 1 + \frac{\partial \alpha_i^{1/2}}{\partial \boldsymbol{\xi}_i} \frac{\boldsymbol{\xi}_i}{\alpha_i^{1/2}} \right| \right) = C - \varphi_i, \quad (8)$$

214 for a constant C and with

$$\varphi_i = [\mathbf{y}^n - \mathbf{H}\mathcal{M}(\mathbf{x}_i^{n-1})]^T (\mathbf{R} + \mathbf{H}\mathbf{Q}\mathbf{H}^T)^{-1} [\mathbf{y}^n - \mathbf{H}\mathcal{M}(\mathbf{x}_i^{n-1})], \quad (9)$$

215 so that

$$p(\mathbf{y}^n | \mathbf{x}_i^{n-1}) \propto e^{-\varphi_i/2}. \quad (10)$$

216 In practice the scale factor α_i is determined numerically by solving

$$\gamma \left(\frac{N_x}{2}, \frac{\alpha_i \boldsymbol{\xi}_i^{n,T} \boldsymbol{\xi}_i^n}{2} \right) = e^{-c_i/2} \gamma \left(\frac{N_x}{2}, \frac{\boldsymbol{\xi}_i^{n,T} \boldsymbol{\xi}_i^n}{2} \right), \quad (11)$$

217 for α_i , where $\gamma(s, x) = \int_0^x t^{s-1} e^{-t} dt$ is the lower incomplete gamma function.
 218 and we refer to $c_i = \max_j [\varphi_j] - \varphi_i$ as the i th offset. See Appendix for details.

219 By using the solution of (11) in the update expression (6) one ensures that
 220 the unnormalized weight associated with the i th updated state vector \mathbf{x}_i^n is equal
 221 to the chosen target weight. The log-weight offsets c_1, c_2, \dots, c_{N_e} are necessary
 222 because the likelihood $p(\mathbf{y}^n | \mathbf{x}_i^{n-1}) \propto \exp(-\varphi_i/2)$ of the current observation
 223 given the previous state of the i th particle will differ between particles. For
 224 every particle i to reach the target weight we need $c_i \geq 0$. That is, the target
 225 unnormalized weight cannot be set larger than the smallest unnormalized weight
 226 in the ensemble. Consequently, since the incomplete γ -function is monotonically
 227 increasing, we must have $\alpha_i \leq 1$ for every particle i . We therefore expect an
 228 updated IEWPF ensemble to have a smaller spread than a sample drawn from
 229 the OPD, and this suggests an explanation for the bias in the original IEWPF.
 230 The offset c_i in equation (11) is chosen by targeting the smallest unnormalized

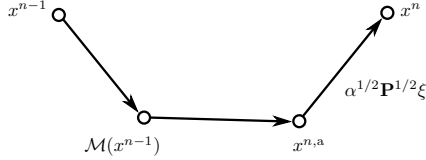


Figure 1: Single-stage IEWPF proposal scheme.

231 weight in the forecast ensemble. In principle, c_i could be defined differently,
 232 targeting for instance the average or median weight. Targeting the smallest
 233 weight, i.e. the maximum φ_i , has the advantage of making all offsets non-
 234 negative, which guarantees that a solution of (11) exists.

235 In their original formulation of the IEWPF, Zhu et al. (2016) considered the
 236 limiting case of (11) when $N_x \rightarrow \infty$. This yields a simplified equation for α_i
 237 which admits an analytical solution in terms of the Lambert W function (Weis-
 238 stein, 2002). A feature of this closed-form solution is a gap between branches
 239 of the Lambert W function, leaving a region in state space where the proposal
 240 density of the filter is zero. The authors used both branches of the solution, one
 241 corresponding to $\alpha_i \leq 1$ and one to $\alpha_i \geq 1$, to reduce the bias of the resulting
 242 filter. We see here that using both solutions is in fact inconsistent, and only
 243 the $\alpha_i \leq 1$ solutions are valid. In this article we do not simplify or approximate
 244 equation (11). Instead we resort to numerical solution methods for determin-
 245 ing α_i . Although the solutions of (11) obtained in this way do not have a gap
 246 between distinct branches, the resulting transformation from ξ_i^n to \mathbf{x}_i^n is not in
 247 general bijective (see Section 2.3).

248 An elementary sketch of the particle movement of the single-stage IEWPF
 249 is summarized in Figure 1. Details of the IEWPF implementation are provided
 250 in the Appendix.

251 **2.3 Properties of the single-stage IEWPF**

252 The implicit formulation of the IEWPF makes it difficult to study the properties
253 of the resulting particle representation. For instance it is not clear, even in sim-
254 plified model settings, how to calculate closed form expressions describing how
255 the IEWPF update changes the ensemble mean or variance. In what follows we
256 will nevertheless gain insight in the solutions via the form of the implicit trans-
257 formation, and by simulating from a Gaussian model where the exact solution
258 is known. In Section 3 we then modify the algorithm and overcome some of the
259 shortcomings of the single-stage IEWPF.

260 Figure 2 shows solutions of the equal weights equation (11) for seven different
261 offsets c (ignoring the subscript i in this display). When implementing the
262 IEWPF, we require $c > 0$, but here we consider the more general case $c \in \mathbb{R}$.
263 The solutions in Figure 2 are shown in terms of the transformation from $\boldsymbol{\xi}$ to
264 $\alpha^{1/2}\boldsymbol{\xi}$. When $c = 0$ the solution is $\alpha = 1$ which gives the identity transformation.
265 Furthermore the solution α decreases with increasing c , so for $c < 0$ the resulting
266 transformation has the effect of expanding the probability distribution of the
267 perturbation $\boldsymbol{\xi}$, whereas for $c > 0$ the transformation contracts the distribution.

268 As can be seen in Figure 2 the contracting solutions for $c > 0$ have horizontal
269 asymptotes while the expanding solutions for $c < 0$ have vertical asymptotes.
270 As a consequence, the transformation from $\boldsymbol{\xi}$ to $\alpha^{1/2}\boldsymbol{\xi}$, and hence to \mathbf{x}^n , is not
271 defined on the whole domain when $c > 0$, and is not surjective when $c < 0$. Only
272 when $c = 0$ is the transformation bijective. When solving (11) with negative
273 offsets therefore, we are not free to use any proposal distribution for $\boldsymbol{\xi}$ as the
274 range of possible perturbation vectors must be restricted to the appropriate
275 subset of the domain. One could try to achieve this by truncating the proposal
276 distribution at the location of the vertical asymptote when $c < 0$. But this
277 is not a viable modification of the IEWPF, because truncation introduces a

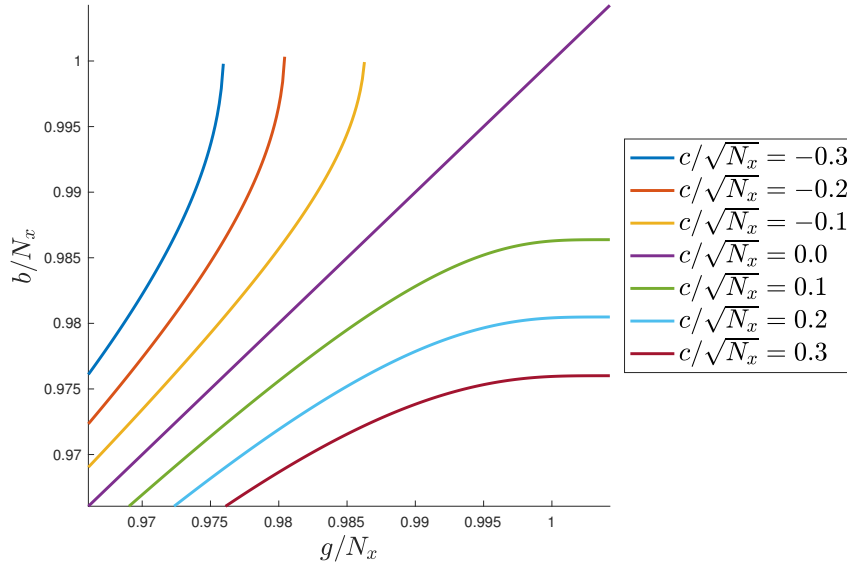


Figure 2: Solutions of (11) for different offsets c shown in terms of $g = \|\xi\|$ and $b = \|\alpha^{1/2}\xi\|$.

278 particle-dependent normalisation constant into the expression for the particle
 279 weights, making them unequal. Here, the requirement that weights be kept
 280 equal appears to be in conflict with the requirement that the transformation
 281 from ξ to \mathbf{x}^n should be a bijection from \mathbb{R}^{N_x} to \mathbb{R}^{N_x} (Chorin et al., 2010). A
 282 theoretical justification of the IEWPF ultimately necessitates the resolution of
 283 this conflict, but it is unclear whether it can be resolved.

284 Considering the OPD and the update expression (6), it is clear that when
 285 $\alpha_i < 1$, the IEWPF produces updated particles with a smaller variance than
 286 the OPD PF, which is known to be unbiased. Hence we expect underestimation
 287 of variance as a consequence of using only contracting solutions of the equal-
 288 weights equation. As is illustrated in the following simulation study, the IEWPF
 289 does indeed tend to underestimate the variability of the state vector in the long
 290 run.

291 To make the presentation of the IEWPF more concrete before introducing
292 the revised version, we now apply the IEWPF to a test case involving a Gauss-
293 linear model. We revisit the same test case in Section 3.2 after describing the
294 revised IEWPF. A more detailed description of the test case is given in Section
295 4.1.

296 We consider a size 100 state vector with initial state $\mathbf{x}^0 \sim N(0, \mathbf{B})$. The
297 transition mechanism is defined by $\mathbf{x}^n \sim N(\mathbf{x}^{n-1}, \mathbf{Q})$, $n = 1, \dots$. Further,
298 observations are given by $\mathbf{y}^n \sim N(\mathbf{x}^n, \mathbf{R})$. The filtering distribution is then
299 Gaussian and its mean and covariance matrix are provided by the Kalman filter
300 (Kalman, 1960). The covariance matrices \mathbf{B} , \mathbf{Q} and \mathbf{R} are all diagonal, with
301 constant diagonal entries of 1.0, 0.04 and 0.12 respectively.

302 When applying the IEWPF as defined by equation (6) and (8) to this model
303 we find that while the ensemble mean matches the KF mean on average, the
304 ensemble spread is too small to match the KF variance in the long run. This
305 means that the IEWPF systematically underestimates the variance of the state
306 (see Figures 3 and 4).

307 In Figure 3 the results of 1000 independent simulations are visualised for one
308 state variable (component 42) at time $n = 120$. We show the rank histogram of
309 the true variable in the set of $N_e = 25$ particles. This is computed by sorting
310 the particles from smallest to largest by the value of this component, and then
311 determining the position of the true value in this ordering. The procedure is
312 repeated for each simulation. When the true state is unavailable the preferred
313 approach is to carry out ranking in data space, comparing observations with
314 realizations of their model equivalents generated from the ensemble. For a
315 detailed treatment of rank histograms, their use and cautions, see Hamill (2001).

316 The rank of the true state relative to the ensemble should ideally be uniform,
317 but in Figure 3 we notice few ranks in the middle. The true value is too often

318 at the extremes of the distribution represented by the 25 particle members.
 319 This means that the ensemble is underdispersive, i.e. the variability in the
 320 particle set is too small. Figure 4 shows the distribution over 1000 simulations
 321 of variance at time $n = 20$, averaged over all 100 entries of the state vector.
 322 A corresponding variance distribution for the stochastic EnKF is included for
 323 comparison. None of the filters being compared use inflation or localization.
 324 The purpose of the comparison is not to show which filter performs better,
 325 but rather to demonstrate that the IEWPF systematically underestimates the
 326 filtering variance. The variance in the particle representation varies somewhat
 327 between the different state vector entries but is mostly between 0.02 and 0.04.
 328 In comparison, the variance calculated by the Kalman filter is 0.052 for all state
 329 vector entries. Both the IEWPF and the EnKF underestimate the long-run
 330 process variability for this example. For the EnKF, variance estimates become
 331 more consistent with the KF level when the ensemble size is increased (Figure
 332 4, bottom display). The same is not true of the IEWPF.

333 In Section 4 we provide further analysis of this example, studying how the
 334 filter behaves over time. We also compare results of the single-stage IEWPF
 335 with our new algorithm using two stages.

336 **3 Modifying the IEWPF**

337 To address the underestimation of variance by the IEWPF described in the
 338 previous section, we now introduce a modified version of the filter. We add
 339 a second perturbation vector $\boldsymbol{\eta}_i$, orthogonal to $\boldsymbol{\xi}_i$, to the analysis state $\mathbf{x}_i^{n,a}$.
 340 We refer to the filter with two separate perturbation vectors as the two-stage
 341 IEWPF, while the original filter with one perturbation vector is referred to
 342 as the single-stage IEWPF. Like the single-stage filter, the two-stage case also
 343 involves a particle-specific parameter α_i which ensures equal particle weights.

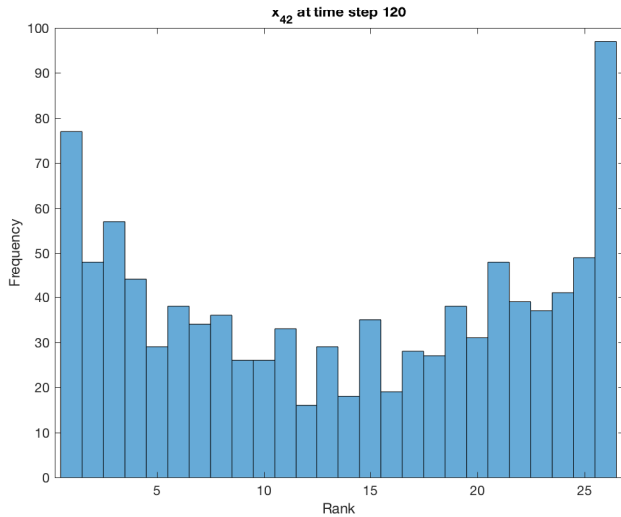


Figure 3: Rank histogram of x_{42}^{120} of true realisation relative to IEWPF ensemble over 1000 simulations. U-shape suggests ensemble is under-dispersed.

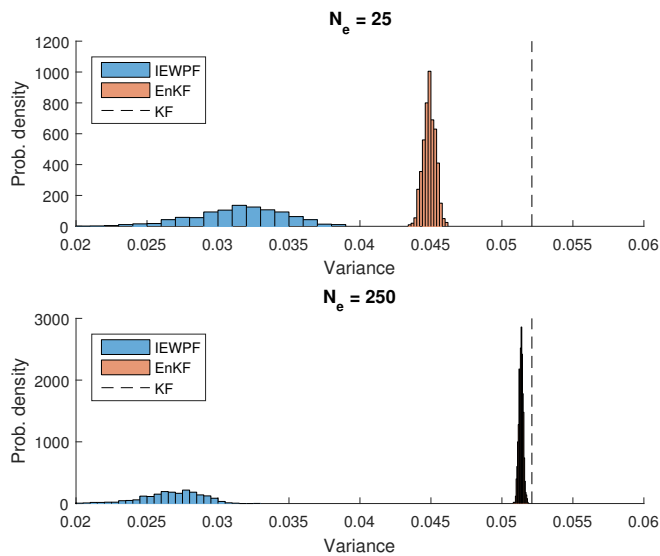


Figure 4: Histograms of estimates of $\text{Var}(\mathbf{x}^{20})$ from 1000 runs of the IEWPF and stochastic EnKF, averaged over $N_x = 100$ elements compared with deterministic KF result. **Top:** $N_e = 25$. **Bottom:** $N_e = 250$.

344 In the two-stage filter there is an additional parameter β , which is common
345 to all particles and is related to the spread of the ensemble. Note that in the
346 single-stage case, α_i depends on the unnormalised weight of the i th particle
347 in the forecast ensemble as well as the magnitude of the sampled perturbation
348 vector $\boldsymbol{\xi}_i$, i.e. $\alpha_{i,1\text{-stage}} = \alpha_i(\varphi_i, \|\boldsymbol{\xi}_i\|)$. In the two-stage case, α_i will also depend
349 on β and the magnitude of $\boldsymbol{\eta}_i$, i.e. $\alpha_{i,2\text{-stage}} = \alpha_i(\varphi_i, \|\boldsymbol{\xi}_i\|, \beta, \|\boldsymbol{\eta}_i\|)$.

350 3.1 Two-stage IEWPF

351 In the two-stage proposal scheme, the updated particle \mathbf{x}_i^n is given by

$$\mathbf{x}_i^n = \mathbf{x}_i^{n,a} + \beta^{1/2} \mathbf{P}^{1/2} \boldsymbol{\eta}_i + \alpha_i^{1/2} \mathbf{P}^{1/2} \boldsymbol{\xi}_i, \quad (12)$$

where the perturbation vectors $\boldsymbol{\xi}_i, \boldsymbol{\eta}_i \in \mathbb{R}^{N_x}$ are standard multivariate Gaussian random vectors satisfying $\boldsymbol{\xi}_i^T \boldsymbol{\eta}_i = 0$. Requiring orthogonality simplifies the particle weight expression so that the equal-weights equation for α_i has the same form as in the single-stage case. Using perturbation vectors that are not orthogonal would introduce extra terms in the equal-weights equation (see Appendix). The equal-weights equation for the updating scheme (12) is

$$\begin{aligned} & (\alpha_i - 1) \boldsymbol{\xi}_i^T \boldsymbol{\xi}_i - 2 \log(\alpha_i^{N_x/2}) - 2 \log \left(\left| 1 + \frac{\partial \alpha_i^{1/2}}{\partial \boldsymbol{\xi}_i} \frac{\boldsymbol{\xi}_i}{\alpha_i^{1/2}} \right| \right) \\ & = C - \varphi_i - (\beta - 1) \boldsymbol{\eta}_i^T \boldsymbol{\eta}_i. \end{aligned} \quad (13)$$

352 Note that (13) is identical to the single-stage equal-weights equation (8) with
353 the offset now defined as $c_i = \max_j [D_j] - D_i$ where $D_j = \varphi_j - (1 - \beta) \boldsymbol{\eta}_j^T \boldsymbol{\eta}_j$.

354 The purpose of the additional perturbation $\boldsymbol{\eta}_i$ and the common scale factor
355 β is to control the spread of the updated particles so that the filter correctly rep-
356 represents the variability of the filtered state. In applications β would be considered
357 a tuning parameter.

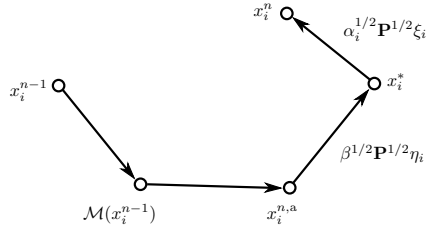


Figure 5: Two-stage IEWPF proposal scheme. Compare with Figure 1.

358 To determine a suitable value of β , particle ranks or coverage probabilities
 359 may be used. Since the spread of the updated ensemble is sensitive to the value
 360 of β , different values will produce differently shaped rank distributions and dif-
 361 ferent observed coverage probabilities. This can indicate whether the currently
 362 used value of β is suitable and, if it is not, whether the value should be ad-
 363 justed up or down. An automated search procedure based on some quantitative
 364 mismatch criterion—say, the difference between observed and nominal coverage
 365 probabilities—is also possible.

366 A coverage probability is the observed frequency with which a prediction
 367 interval covers the predicted quantity. Ideally it should match the interval's
 368 nominal confidence level. For instance, an 80% prediction interval for y^n is
 369 $(y_{(0.1N_e)}^n, y_{(0.9N_e)}^n)$, and on average about 80% of the data vector entries at time
 370 n should fall within this interval. We suggest tuning β such that the coverage
 371 probabilities at the 50%, 60%, \dots , 90% levels all match their respective nominal
 372 confidence level reasonably well. This entails running the algorithm for a range
 373 of β values, and choosing a value that gives an acceptable calibration (see Section
 374 4 for more details about how this is tuned in practical experiments).

375 An elementary sketch of the particle movement of the two-stage IEWPF is
 376 summarized in Figure 5. Implementation details are provided in the Appendix.

377 3.2 Properties of the two-stage IEWPF

378 As for the single-stage IEWPF, it is difficult to study analytical properties of
379 the two-stage IEWPF, even in simplified model settings. Some insight can still
380 be gleaned by simulating from a Gaussian model where the exact solution is
381 known.

382 In section 2.3 we stated that the single-stage transformation from $\boldsymbol{\xi}$ to \mathbf{x}
383 implied by (6) is only injective for $c \geq 0$ and is only surjective for $c \leq 0$, i.e. it
384 is a bijection only when $c = 0$. In the two-stage case we can think of the map
385 from $\boldsymbol{\xi}$ to \mathbf{x} as depending on $\boldsymbol{\eta}$ and β through c . That is, there is not one map
386 $\psi : \boldsymbol{\xi} \mapsto \mathbf{x}$, but a set $\{\psi_c : c \geq 0\}$ of maps where c is a function of β and $\boldsymbol{\eta}$. The
387 two-stage IEWPF keeps β fixed and draws a random $\boldsymbol{\eta}$, thereby selecting one
388 of the maps ψ_c . Then $\boldsymbol{\xi}$ is drawn subject to the orthogonality constraint. For
389 any point $\mathbf{x} \in \mathbb{R}^{N_x}$ and any $c \geq 0$, there is some combination of $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ with
390 $\boldsymbol{\eta}^T \boldsymbol{\xi} = 0$ such that ψ_c maps $\boldsymbol{\xi}$ onto \mathbf{x} . With $\boldsymbol{\eta}$ fixed, there may not exist a $\boldsymbol{\xi}$ that
391 is orthogonal to $\boldsymbol{\eta}$ and is mapped onto \mathbf{x} . Introducing a second perturbation
392 vector to randomize the selection of a map is thus a way to ensure that state
393 space is covered by the proposal distribution.

394 Since the proposal distribution of the additional perturbation vector is zero-
395 mean, the expectation of the state vector is the same under the two-stage update
396 scheme as under the single-stage scheme. Hence, the modification does not
397 induce a bias in the ensemble mean.

398 We return now to the Gauss-linear model from Section 2.3. A more detailed
399 description of the test case is given Section 4.1. This time we apply the two-
400 stage IEWPF to the Gauss-linear test case. Results of this method and that of
401 the Kalman filter are shown in Figures 6 and 7. As in Section 2.3, these are the
402 results of 1000 independent simulations, and the results are presented for time
403 $n = 120$. Also as in Section 2.3, the ensemble size is $N_e = 25$.

404 Figure 6 shows the rank histograms for the true value of state vector entry
405 42. The rank histogram for $\beta = 0.05$ is clearly U-shaped. As β increases to
406 0.25 and 0.30 the rank distribution becomes more uniform. The rank histogram
407 for $\beta = 0.5$ is indistinguishable from a uniform distribution given the sampling
408 error and the calibration is better than when using the single-stage approach as
409 shown in Figure 3.

410 Figure 7 shows the distributions of average variances produced by the two-
411 stage IEWPF for β set to 0.05, 0.25, 0.30 and 0.50. The average is taken over all
412 elements of the state vector. The display also shows the Kalman filter variance
413 estimate as a thin, vertical line. Ideally the IEWPF should produce an ensemble
414 whose variance matches the KF variance. Of the four β -values considered,
415 0.3 and 0.5 come closest to realizing this, showing a clear improvement over
416 the variance distribution of the single-stage IEWPF in Figure 4. Judging by
417 Figure 7, the optimal value of β in terms of variance calibration seems to lie
418 closer to 0.3 than to 0.5 in this case. Yet Figure 6 shows a more uniform
419 rank distribution for $\beta = 0.5$ than for $\beta = 0.3$. It is important to keep in
420 mind, however, that comparing the rank histograms in terms of their degree of
421 departure from uniformity is less precise than comparing the more concentrated
422 variance histograms in terms of their locations along the horizontal axis. Figure
423 7 is therefore probably a better guide to identifying the optimal value of β .
424 On the other hand, it cannot be ruled out that the discrepancy between the
425 two figures has a different cause, such as the updated particles having a non-
426 Gaussian distribution.

427 4 Numerical experiments

428 We present two synthetic test cases for assessing the performance of the IEWPF
429 algorithms described in sections 2 and 3. The first is a Gauss-linear test case

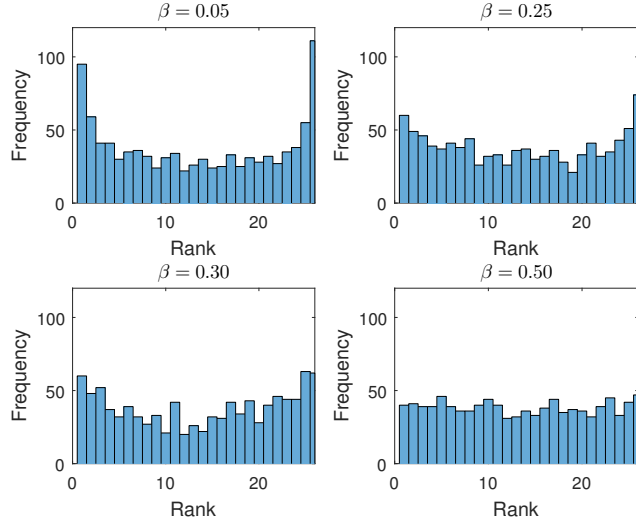


Figure 6: Rank histogram of x_{42}^{120} of true realisation relative to IEWPF ensemble over 1000 simulations. Results are for the two-stage IEWPF using four different values of β .

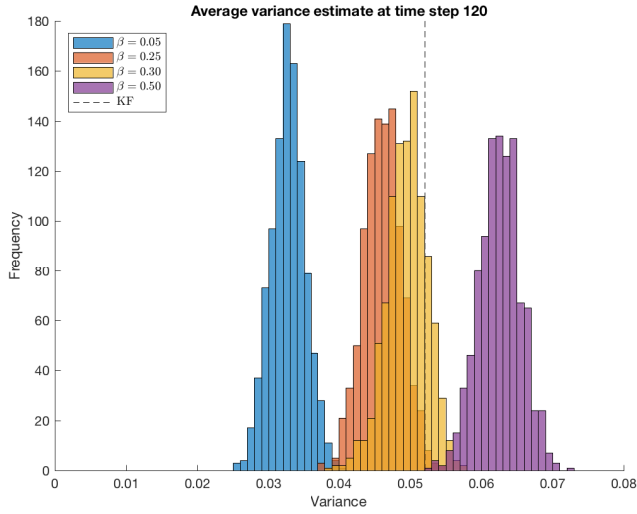


Figure 7: Histograms of two-stage IEWPF estimates of the variance of \mathbf{x}^{120} , averaged over all elements, for four different values of β , based on 1000 simulations each.

430 where the dynamical system evolves according to a linear model and the system
 431 state is observed directly except for an additive observation error term. In the
 432 second case the state evolves according to the Lorenz96 model (Lorenz, 1995)
 433 and we observe every second element of the state vector. We assume Gaussian
 434 probability distributions for the initial state, model errors and observation errors
 435 as described in section 2.1. In the Gauss-linear case the filtering distribution
 436 is analytically available via the Kalman filter under these assumptions, and we
 437 will make use of this to judge the quality of the estimates produced by the
 438 single-stage and two-stage IEWPFs.

439 4.1 Gauss-linear model

This is the Gauss-linear test case referred to in Sections 2.3 and 3.2. We re-use
 the model and observation equations from Section 3.1 of Zhu et al. (2016):

$$\mathbf{x}^n = \mathbf{x}^{n-1} + \mathbf{u}^n, \quad (14)$$

$$\mathbf{y}^n = \mathbf{x}^n + \mathbf{v}^n, \quad (15)$$

$$\mathbf{u}^n \sim N(0, \mathbf{Q}), \quad \mathbf{v}^n \sim N(0, \mathbf{R}), \quad \mathbf{x}^0 \sim N(0, \mathbf{B}),$$

$$N_x = 100, \quad n_t = 120, \quad \mathbf{Q} = 0.04\mathbf{I}, \quad \mathbf{R} = 0.12\mathbf{I}, \quad \mathbf{B} = \mathbf{I}.$$

The filtering probability density $p(\mathbf{x}^n | \mathbf{y}^1, \dots, \mathbf{y}^n)$ is Gaussian with param-

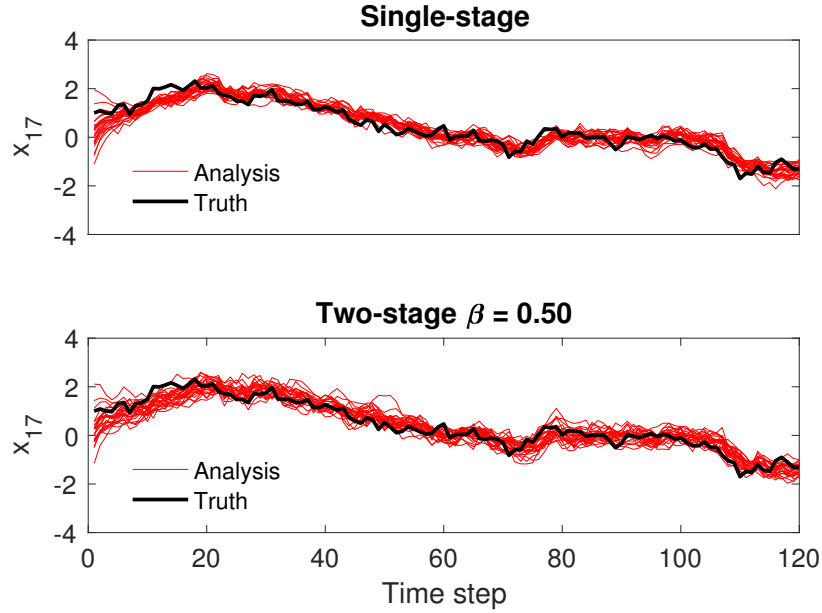


Figure 8: Example trajectories of x_{17} under Gauss-linear model. **Top:** Single stage IEWPF. **Bottom:** Two-stage IEWPF with $\beta = 0.5$. In both panels analysis ensemble members are shown in red and the true model trajectory in black.

ters $\boldsymbol{\mu}^n$ and \mathbf{P}^n , given recursively via the Kalman filter:

$$\begin{aligned}\boldsymbol{\mu}^n &= \boldsymbol{\mu}^{n-1} + (\mathbf{P}^{n-1} + \mathbf{Q})(\mathbf{R} + \mathbf{P}^{n-1} + \mathbf{Q})^{-1}(\mathbf{y}^n - \boldsymbol{\mu}^{n-1}) \\ \mathbf{P}^n &= \mathbf{P}^{n-1} + \mathbf{Q} - (\mathbf{P}^{n-1} + \mathbf{Q})(\mathbf{R} + \mathbf{P}^{n-1} + \mathbf{Q})^{-1}(\mathbf{P}^{n-1} + \mathbf{Q}),\end{aligned}\quad (16)$$

440 where $\boldsymbol{\mu}^0 = \mathbf{0}$ and $\mathbf{P}^0 = \mathbf{B}$.

441 We compare results of the single-stage and two-stage IEWPF, using the KF
442 filtering distribution (16) as a reference solution. The number of particles is
443 $N_e = 25$, and we run the algorithm for 1000 simulations.

444 Example trajectories of the single-stage and two-stage IEWPF algorithms
445 are shown in Figure 8. Both follow the true state pretty well, but the single-stage
446 results (top display) have less variability.

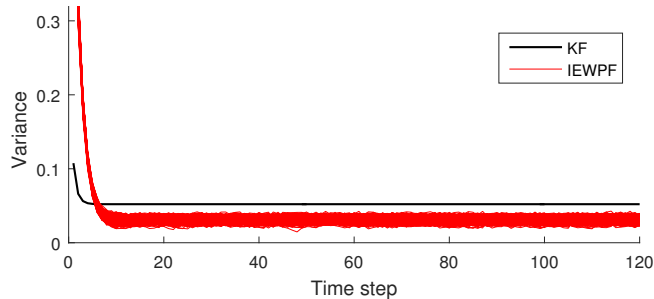


Figure 9: Comparison of posterior variance estimates from the Kalman filter and the single-stage IEWPF for state vector entry 42 in the Gaussian linear model case with $N_x = 100$ and $N_e = 25$. The IEWPF variance curves show the result of 1000 repetitions of the filtering task.

447 Figure 9 shows variance results for the single-stage IEWPF over the assim-
 448 ilation steps. Initial filtering variance, specified through the background error
 449 covariance matrix \mathbf{B} , is 1. The variance of the KF filtering distribution decreases
 450 quickly before stabilising, while the IEWPF ensemble variance takes longer to
 451 stabilise, and does so at a lower variance level. Comparing the IEWPF and
 452 KF variance estimates, it is clear that the IEWPF overestimates the filtering
 453 variance early on, and underestimates it in the long run.

454 Figure 10 shows average two-stage IEWPF variance estimates over the as-
 455 simulation steps. As in the single-stage case (Figure 9), variability is still over-
 456 estimated at the beginning of the time interval, but the long-run KF variance
 457 can now be matched quite well by an appropriate choice of β .

458 4.2 Lorenz96 model

459 We study the performance of the single and two-stage IEWPF using the model
 460 presented in Section 3.2 of Zhu et al. (2016). The dynamical model is given by

$$\frac{dx_i}{dt} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F, \quad i = 1, \dots, N_x, \quad (17)$$

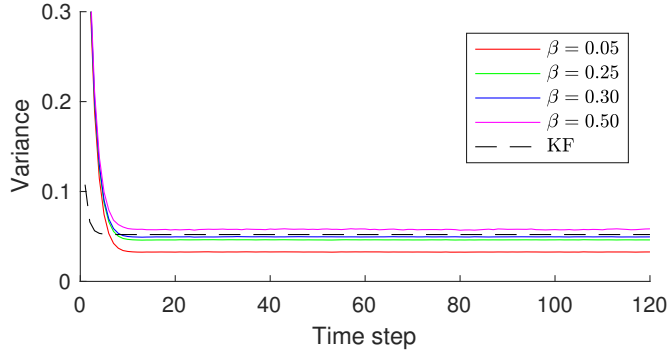


Figure 10: Evolution of variance estimates from the two-stage IEWPF with four different values of the common scale factor β . The solid curves show variance estimates averaged over 1000 independent simulations. The Kalman filter variance estimate is included for comparison.

461 where the indices wrap around so that x_{N_x+1} is identified with x_1 . Letting
 462 $\mathbf{x}^n = \mathbf{x}(t_n)$ with $t_n = n\Delta t$, the model equation can be written as

$$\mathbf{x}^n = \mathcal{M}(x^{n-1}) + \mathbf{u}^n, \quad \mathbf{u}^n \sim N(0, \mathbf{Q}), \quad n = 1, \dots, n_t,$$

463 where \mathcal{M} denotes integration of equation (17) by a fourth order Runge-Kutta
 464 scheme and

$$N_x = 40, \quad n_t = 300, \quad F = 8, \quad \Delta t = 0.05.$$

465 Observations are gathered at every time step t_n , $n = 1, \dots, n_t$, which means
 466 that here Δt is both the integration time step of the numerical solution of (17)
 467 and the time between successive observation time points. Data are related to
 468 the state vector by

$$\mathbf{y}^n = \mathbf{H}\mathbf{x}^n + \mathbf{v}^n, \quad \mathbf{v}^n \sim N(0, \mathbf{R}),$$

469 where \mathbf{H} is a selection matrix which picks out every second element of the state
 470 vector, so that $\mathbf{H}\mathbf{x}^n = (x_2^n, x_4^n, \dots, x_{N_x}^n)^T$. The remaining model parameters
 471 are specified as follows:

$$\mathbf{B} = \text{tridiag}(0.25, 1, 0.25), \quad \mathbf{Q} = \text{tridiag}(0.025, 0.10, 0.025), \quad \mathbf{R} = 0.16I,$$

472 where $\text{tridiag}(a_1, a_2, a_3)$ is a tridiagonal matrix with a_1 in every entry of the first
 473 subdiagonal, a_2 on the main diagonal and a_3 on the first superdiagonal. The
 474 nonlinearity in this data assimilation test case is weak due to the high frequency
 475 of observations. Increasing the time between updates would give a more severe
 476 test of the filter. A weakly nonlinear test case is still suitable for demonstrating
 477 that the IEWPF ensemble spread can be controlled through the choice of β .

478 We run the single-stage and two-stage IEWPF variants on this test case,
 479 with $N_e = 100$ in both cases. Figure 11 shows results of the two-stage IEWPF
 480 using $\beta = 0.7$, where we plot the filtering distribution over time along with the
 481 truth. This is done for two entries of the state vector (component 1 and 2). The
 482 ensemble tracks the reference state and covers the truth reasonably well. The
 483 bottom display shows the estimated variances of component 1 and 2. Because
 484 the observations provide much more information about the second entry, this
 485 has smaller variance over time.

486 In Figure 12 we plot coverage probabilities at one time step. These are
 487 plotted for different β parameters and for different confidence levels. The tuning
 488 procedure tells us that a value of β near 1 is useful in this example because it
 489 gives the best predictive performance, and any value in the range $0.7 - 1.2$
 490 would be acceptable. Figure 13 compares rank histograms of one run each of
 491 the single-stage IEWPF and the two-stage IEWPF with $\beta = 0.7$, the latter
 492 being the same run used to make Figure 11. The single-stage rank histogram
 493 has a clear U-shape while the two-stage rank histogram does not, suggesting

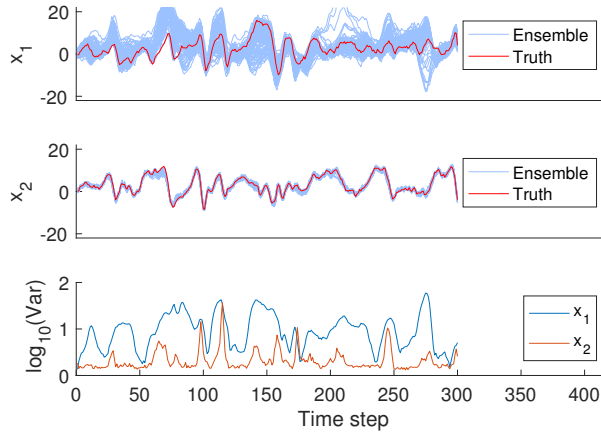


Figure 11: Time evolution of true realisations and an IEWPF ensemble. The components shown are x_1 (not directly observed) and x_2 (directly observed) along with ensemble variances. This ensemble was obtained from the two-stage IEWPF with $\beta = 0.7$ applied to the standard Lorenz96 case with $N_x = 40$.

494 that the two-stage IEWPF is better calibrated.

495 Finally, to test the two-stage IEWPF in a setting that is both weakly non-
496 linear and where N_x is much larger than N_e , we run the Lorenz96 case with
497 $N_x = 1000$, $N_e = 25$ and $\beta = 0.75$. Remaining parameter values are unchanged.
498 Figure 14 shows two components of the estimated and true model trajectories in
499 this moderately high-dimensional test case. The top and bottom displays show
500 particle trajectories for an observed component and an unobserved component
501 of the state vector respectively. Filter behaviour is not noticeably different from
502 the lower dimensional case of Figure 11. The variance is clearly larger for the
503 unobserved state. For both variables, coverage is reasonable, and no bias effects
504 are apparent. As is common in Lorenz models, the state is sometimes very un-
505 certain; for instance at time steps 120–150, and especially so for the unobserved
506 state. Even though the state dimension is much larger here, it seems that β
507 can be in the same range ($\beta = 0.75$ in this plot as opposed to $\beta = 0.7$ in the
508 $N_x = 40$ case).

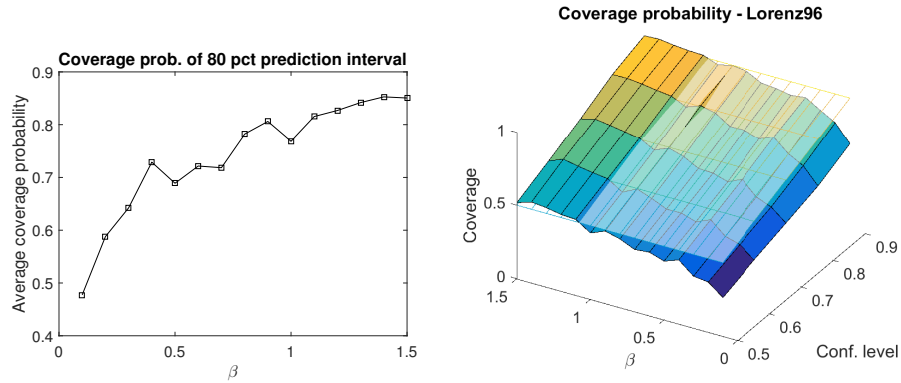


Figure 12: Empirical coverage probability as a function of prediction interval confidence and the value of the tuning parameter β . Best calibration is achieved for $\beta \approx 1$. **Left:** Average empirical coverage probability of 80 per cent prediction intervals constructed from two-stage IEWPF ensembles with a range of β values. **Right:** Coverage probabilities as a function of β for a range of confidence levels.

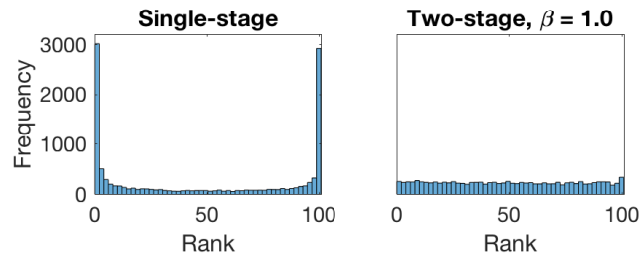


Figure 13: Rank histograms for one run of the single-stage IEWPF and one run of the two-stage IEWPF with $\beta = 0.7$ on the Lorenz96 model test case. Ranks are aggregated over all steady-state time steps and all state elements.

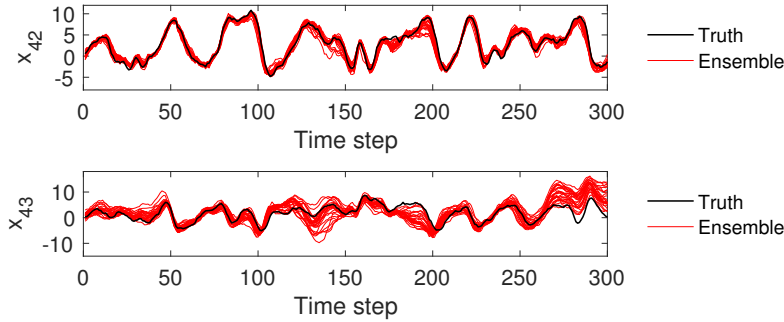


Figure 14: Two components of the trajectories of the ground truth and two-stage IEWPF ensemble in the high-dimensional Lorenz96 case with $N_x = 1000$ and $N_e = 25$. The components shown are x_{42} (directly observed) and x_{43} (not directly observed). The common scale factor in the two-stage IEWPF was set to $\beta = 0.75$ in this case.

509 5 Conclusion

510 In this paper we have presented a modification to the Implicit Equal Weight
 511 Particle Filter (IEWPF). The suggested approach is applicable to data assimilation
 512 in both low and high-dimensional state space models. When applied to
 513 a weakly nonlinear dynamical model, the revised IEWPF performed reasonably
 514 well even in the $N_x \gg N_e$ case.

515 A MATLAB implementation of the IEWPF algorithm ran in seconds to min-
 516 utes when applied to the Gauss-linear and Lorenz96 data assimilation test cases
 517 in Section 4. The runtime increases in proportion with the number of time
 518 steps. It is not sensitive to the state dimension, but the numerical solution of
 519 the equal-weights equation may require more iterations to achieve convergence
 520 for very high-dimensional cases. The IEWPF update can be carried out in par-
 521 allel for each particle once the weight-offsets for the whole ensemble have been
 522 determined, making the algorithm easy to parallelize. The memory require-
 523 ments of the IEWPF are largely determined by the need to store the model
 524 error and observation error covariance matrices.

525 When using the revised filtering method, particles are updated according
526 to a two-stage proposal scheme which draws two separate and orthogonal per-
527 turbation vectors from the proposal distribution. By using two stages, we are
528 able to eliminate the gap in the proposal distribution of the original IEWPF,
529 ensuring that the proposal distribution is nonzero everywhere in state space.
530 The random perturbations are scaled to keep weights equal and to achieve the
531 correct ensemble spread. Accurately adjusting the spread requires tuning of
532 the corresponding scale parameter. In our setting this is a single parameter
533 which we propose to specify through the use of coverage probabilities or rank
534 histograms. Other approaches might be possible here, for instance a criterion
535 guided by the distribution of weights in the optimal particle density proposal.
536 We leave this for future work. Note that we have chosen to keep the tuning
537 parameter fixed throughout the data assimilation period. One could also adjust
538 this parameter dynamically, so that different values can be used at different
539 assimilation times as is done in adaptive inflation for EnKFs.

540 The updating schemes described in this paper are constructed so that the
541 distribution of particle weights will be uniform. An alternative would be to fix
542 some proportion of the weights, allowing the rest to vary. This has the possible
543 benefit of balancing contracting and expanding solutions of the equal-weights
544 equation. Another possibility is to select multiple target weights, so that the
545 overall distribution of weights is uniform within sub-ensembles of particles, but
546 may differ between sub-ensembles.

547 **References**

548 Ades, M. and Van Leeuwen, P. J. (2013). An exploration of the equivalent
549 weights particle filter. *Quarterly Journal of the Royal Meteorological Society*,
550 139(672):820–840.

- 551 Asch, M., Bocquet, M., and Nodet, M. (2016). *Data assimilation: methods,*
552 *algorithms, and applications*, volume 11. SIAM, Philadelphia.
- 553 Brookes, M. (2011). The matrix reference manual. *URL*
554 *<http://www.ee.imperial.ac.uk>*.
- 555 Chorin, A., Morzfeld, M., and Tu, X. (2010). Implicit particle filters for data
556 assimilation. *Communications in Applied Mathematics and Computational*
557 *Science*, 5(2):221–240.
- 558 Chorin, A. J., Morzfeld, M., and Tu, X. (2013). A survey of implicit particle
559 filters for data assimilation. In *State-Space Models*, pages 63–88. Springer.
- 560 Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential monte carlo
561 sampling methods for bayesian filtering. *Statistics and computing*, 10(3):197–
562 208.
- 563 Evensen, G. (2009). *Data assimilation: the ensemble Kalman filter*. Springer
564 Science & Business Media.
- 565 Fletcher, S. J. (2017). *Data Assimilation for the Geosciences: From Theory to*
566 *Application*. Elsevier, Amsterdam.
- 567 Frei, M. and Künsch, H. R. (2013). Bridging the ensemble kalman and particle
568 filters. *Biometrika*, 100(4):781–800.
- 569 Gordon, N. J., Salmond, D. J., and Smith, A. F. (1993). Novel approach
570 to nonlinear/non-gaussian bayesian state estimation. In *IEE Proceedings F*
571 *(Radar and Signal Processing)*, volume 140(2), pages 107–113. IET.
- 572 Hamill, T. M. (2001). Interpretation of rank histograms for verifying ensemble
573 forecasts. *Monthly Weather Review*, 129(3):550–560.

- 574 Kalman, R. E. (1960). A new approach to linear filtering and prediction prob-
575 lems. *Journal of basic Engineering*, 82(1):35–45.
- 576 Lorenz, E. N. (1995). Predictability: A problem partly solved, vol. i. In *Pro-*
577 *ceedings of Seminar at ECMWF*.
- 578 Morzfeld, M., Tu, X., Atkins, E., and Chorin, A. J. (2012). A random
579 map implementation of implicit filters. *Journal of Computational Physics*,
580 231(4):2049–2066.
- 581 Rezaie, J. and Eidsvik, J. (2012). Shrunked $(1 - \alpha)$ ensemble kalman filter and α
582 gaussian mixture filter. *Computational Geosciences*, 16(3):837–852.
- 583 Snyder, C., Bengtsson, T., Bickel, P., and Anderson, J. (2008). Obstacles to
584 high-dimensional particle filtering. *Monthly Weather Review*, 136(12):4629–
585 4640.
- 586 Snyder, C., Bengtsson, T., and Morzfeld, M. (2015). Performance bounds
587 for particle filters using the optimal proposal. *Monthly Weather Review*,
588 143(11):4750–4761.
- 589 Stordal, A. S., Karlsen, H. A., Nævdal, G., Skaug, H. J., and Vallès, B. (2011).
590 Bridging the ensemble kalman filter and particle filters: the adaptive gaussian
591 mixture filter. *Computational Geosciences*, 15(2):293–305.
- 592 Van Leeuwen, P. J. (2009). Particle filtering in geophysical systems. *Monthly*
593 *Weather Review*, 137(12):4089–4114.
- 594 Van Leeuwen, P. J. (2010). Nonlinear data assimilation in geosciences: an ex-
595 tremely efficient particle filter. *Quarterly Journal of the Royal Meteorological*
596 *Society*, 136(653):1991–1999.
- 597 Van Leeuwen, P. J., Cheng, Y., and Reich, S. (2015). *Nonlinear data assimila-*
598 *tion*. Springer.

599 Weisstein, E. W. (2002). Lambert W-function. From MathWorld—A Wolfram
600 Web Resource. <http://mathworld.wolfram.com/LambertW-Function.html>
601 Last accessed on 1 August 2018.

602 Zhu, M., Van Leeuwen, P. J., and Amezcua, J. (2016). Implicit equal-
603 weights particle filter. *Quarterly Journal of the Royal Meteorological Society*,
604 142(698):1904–1919.

605 Appendix

606 Single-stage IEWPF equal-weights equation

607 Suppose we have a forecast ensemble $\{\mathcal{M}(\mathbf{x}_i^{n-1})\}_{i=1}^{N_e}$ with equal weights at the
608 previous time step, and we want to update this ensemble with respect to the
609 observation \mathbf{y}^n . We sample the perturbation vector $\boldsymbol{\xi}_i^n$ from the proposal dis-
610 tribution $q(\boldsymbol{\xi}_i^n)$. The updated weight of particle i is

$$w_i^n = \frac{p(\mathbf{x}_i^n | \mathbf{x}_i^{n-1}, \mathbf{y}^n) p(\mathbf{y}^n | \mathbf{x}_i^{n-1})}{q(\boldsymbol{\xi}_i^n)} \left\| \frac{\partial \mathbf{x}_i^n}{\partial \boldsymbol{\xi}_i^n} \right\|$$

611 Taking $-2 \log$ of both sides gives

$$-2 \log w_i^n = -2 \log p(\mathbf{x}_i^n | \mathbf{x}_i^{n-1}, \mathbf{y}^n) - 2 \log p(\mathbf{y}^n | \mathbf{x}_i^{n-1}) + 2 \log q(\boldsymbol{\xi}_i^n) - 2 \log \left(\left\| \frac{\partial \mathbf{x}_i^n}{\partial \boldsymbol{\xi}_i^n} \right\| \right)$$

and using that $p(\mathbf{x}_i^n | \mathbf{x}_i^{n-1}, \mathbf{y}^n) \propto \exp(-\frac{1}{2}(\mathbf{x}_i^n - \mathbf{x}_i^{n,a})^T \mathbf{P}^{-1}(\mathbf{x}_i^n - \mathbf{x}_i^{n,a}))$ and $p(\mathbf{y}^n | \mathbf{x}_i^{n-1}) \propto \exp(-\frac{1}{2}\varphi_i)$ this becomes

$$\begin{aligned}
-2 \log w_i^n &= (\mathbf{x}_i^n - \mathbf{x}_i^{n,a})^T \mathbf{P}^{-1}(\mathbf{x}_i^n - \mathbf{x}_i^{n,a}) + \varphi_i - (\boldsymbol{\xi}_i^n)^T \boldsymbol{\xi}_i^n - 2 \log \left(\left\| \frac{\partial \mathbf{x}_i^n}{\partial \boldsymbol{\xi}_i^n} \right\| \right) \\
&= (\alpha_i - 1)(\boldsymbol{\xi}_i^n)^T \boldsymbol{\xi}_i^n + \varphi_i^n - 2 \log \left(\left\| \frac{\partial \mathbf{x}_i^n}{\partial \boldsymbol{\xi}_i^n} \right\| \right) \\
&= (\alpha_i - 1)(\boldsymbol{\xi}_i^n)^T \boldsymbol{\xi}_i^n + \varphi_i^n - 2N_x \log \alpha_i^{1/2} \\
&\quad - 2 \log \left(\left| 1 + \frac{\partial \alpha_i^{1/2}}{\partial \boldsymbol{\xi}_i^n} \frac{\boldsymbol{\xi}_i^n}{\alpha_i^{1/2}} \right| \right) - 2 \log \|\mathbf{P}^{1/2}\|
\end{aligned}$$

612 where we have used the single-stage IEWPF update scheme (6) and rewritten
613 the determinant of the Jacobian using Sylvester's determinant lemma (Brookes,
614 2011). Equating the negative log-weight with a constant C now gives

$$(\alpha_i - 1)(\boldsymbol{\xi}_i^n)^T \boldsymbol{\xi}_i^n + \varphi_i^n - 2N_x \log \alpha_i^{1/2} - 2 \log \left(\left| 1 + \frac{\partial \alpha_i^{1/2}}{\partial \boldsymbol{\xi}_i^n} \frac{\boldsymbol{\xi}_i^n}{\alpha_i^{1/2}} \right| \right) - 2 \log \|\mathbf{P}^{1/2}\| = C$$

615 OR

$$(\alpha_i - 1)(\boldsymbol{\xi}_i^n)^T \boldsymbol{\xi}_i^n - 2N_x \log \alpha_i^{1/2} - 2 \log \left(\left| 1 + \frac{\partial \alpha_i^{1/2}}{\partial \boldsymbol{\xi}_i^n} \frac{\boldsymbol{\xi}_i^n}{\alpha_i^{1/2}} \right| \right) = c_i \quad (18)$$

616 where $c_i = C - \varphi_i$ and $2 \log \|\mathbf{P}^{1/2}\|$ has been absorbed into C since it is the
617 same for all particles. If it is assumed that α_i depends on $\boldsymbol{\xi}_i^n$ only through
618 $g_i = (\boldsymbol{\xi}_i^n)^T \boldsymbol{\xi}_i^n$, then the above equation simplifies to

$$(\alpha_i - 1)g_i - 2 \log \left(\alpha_i^{N_x/2-1} \frac{\partial \alpha_i g_i}{\partial g_i} \right) = c_i.$$

619 Writing $b_i = \alpha_i g_i$, this is

$$b_i - g_i + 2 \log g_i^{N_x/2-1} - 2 \log \left(b_i^{N_x/2-1} \left| \frac{\partial b_i}{\partial g_i} \right| \right) = c_i.$$

Separating the terms involving b_i and g_i gives

$$\begin{aligned} \log \left(\exp \left(\frac{b_i}{2} \right) b_i^{N_x/2} \left| \frac{\partial b_i}{\partial g_i} \right| \right) &= \log \left(\exp \left(\frac{g_i}{2} \right) g_i^{N_x/2-1} \right) - \frac{c_i}{2} \\ \exp \left(-\frac{b_i}{2} \right) b_i^{N_x/2-1} \left| \frac{\partial b_i}{\partial g_i} \right| &= \exp \left(-\frac{g_i}{2} \right) g_i^{N_x/2-1} \exp \left(-\frac{c_i}{2} \right) \end{aligned}$$

620 which, when integrated from $g_i = 0$ to $g_i = \tilde{g}_i$, yields the single-stage equal-
621 weights equation (11).

622 Two-stage IEWPF equal-weights equation

623 In the two-stage IEWPF we draw two orthogonal perturbation vectors $\boldsymbol{\xi}_i^n$ and $\boldsymbol{\eta}_i^n$
624 from the proposal distribution $q(\boldsymbol{\xi}_i^n, \boldsymbol{\eta}_i^n)$, and use them to compute the updated
625 particle position according to the two-stage update scheme (12).

626 Orthogonal pairs of multivariate normal perturbation vectors are generated
627 as follows:

- 628 1. Generate $\boldsymbol{\eta}$ and \mathbf{z} by sampling from the standard N_x -variate Gaussian
629 distribution.
- 630 2. Decompose \mathbf{z} into two components $\mathbf{z} = \mathbf{z}_{\parallel} + \mathbf{z}_{\perp}$ where \mathbf{z}_{\parallel} is parallel to $\boldsymbol{\eta}$
631 and \mathbf{z}_{\perp} is orthogonal to $\boldsymbol{\eta}$.
- 632 3. Let $\boldsymbol{\xi} = \sqrt{(\mathbf{z}^T \mathbf{z}) / (\mathbf{z}_{\perp}^T \mathbf{z}_{\perp})} \mathbf{z}_{\perp}$, so that $\boldsymbol{\xi}^T \boldsymbol{\xi} = \mathbf{z}^T \mathbf{z}$.

633 For a pair $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathbb{R}^{N_x}$ satisfying $\boldsymbol{\xi}^T \boldsymbol{\eta} = 0$, we have

$$q(\boldsymbol{\xi}, \boldsymbol{\eta}) \propto \exp \left(-\frac{1}{2} \boldsymbol{\eta}^T \boldsymbol{\eta} \right) \exp \left(-\frac{1}{2} \boldsymbol{\xi}^T \boldsymbol{\xi} \right) I(\boldsymbol{\xi}^T \boldsymbol{\eta} = 0), \quad (19)$$

634 where $I(\boldsymbol{\xi}^T \boldsymbol{\eta} = 0)$ is an indicator function that is equal to one if $\boldsymbol{\xi}^T \boldsymbol{\eta} = 0$ and
635 is equal to zero if $\boldsymbol{\xi}^T \boldsymbol{\eta} \neq 0$.

636 Under the two-stage scheme, assuming that since β is shared between parti-
637 cles the Jacobian matrix of the map from $\boldsymbol{\eta}_i^n$ to \mathbf{x}_i^n can be omitted, the expression

638 for the weight of particle i is

$$w_i^n = \frac{p(\mathbf{x}_i^n | \mathbf{x}_i^{n-1}, \mathbf{y}^n) p(\mathbf{y}^n | \mathbf{x}_i^{n-1})}{q(\boldsymbol{\xi}_i^n, \boldsymbol{\eta}_i^n)} \left\| \frac{\partial \mathbf{x}_i^n}{\partial \boldsymbol{\xi}_i^n} \right\|$$

and taking $-2 \log$ and proceeding as in the single-stage case now gives

$$\begin{aligned} & (\alpha_i - 1)(\boldsymbol{\xi}_i^n)^T \boldsymbol{\xi}_i^n + (\beta - 1)(\boldsymbol{\eta}_i^n)^T \boldsymbol{\eta}_i^n + \varphi_i^n - 2N_x \log \alpha_i^{1/2} \\ & - 2 \log \left(\left| 1 + \frac{\partial \alpha_i^{1/2}}{\partial \boldsymbol{\xi}_i^n} \frac{\boldsymbol{\xi}_i^n}{\alpha_i^{1/2}} \right| \right) - 2 \log \|\mathbf{P}^{1/2}\| = C \end{aligned}$$

639 and

$$(\alpha_i - 1)(\boldsymbol{\xi}_i^n)^T \boldsymbol{\xi}_i^n - 2N_x \log \alpha_i^{1/2} - 2 \log \left(\left| 1 + \frac{\partial \alpha_i^{1/2}}{\partial \boldsymbol{\xi}_i^n} \frac{\boldsymbol{\xi}_i^n}{\alpha_i^{1/2}} \right| \right) = C - \varphi_i - (\beta - 1)(\boldsymbol{\eta}_i^n)^T \boldsymbol{\eta}_i^n$$

640 which matches equation (18) if we let

$$c_i = C - \varphi_i - (\beta - 1)(\boldsymbol{\eta}_i^n)^T \boldsymbol{\eta}_i^n = C + (1 - \beta)(\boldsymbol{\eta}_i^n)^T \boldsymbol{\eta}_i^n - \varphi_i. \quad (20)$$

641 Consequently α_i can be determined in the two-stage case by solving equation

642 (11) with offset c_i given by (20).