1. $\hat{\sigma}^{(\cdot t+1)}(X^{(t+1)}) = \sum_{X^{(t)}} P(X^{(t+1)} | X^{(t)}, H^{(t)})\, \hat{\sigma}^{(t)}(X^{(t)})$

   $\hat{\sigma}^{(t+1)}(X^{(t+1)}) = \hat{\sigma}^{(\cdot t+1)}(X^{(t+1)}) \prod_{k=1}^{K} \sum_{M_k} P(D_k^{(t+1)} | X^{(t+1)}, M_k)\, \hat{\sigma}^{(t)}(M_k)$
   (before normalization)

$$\hat{\sigma}^{(t+1)}(M_k) = \sum_{X^{(t)}} P(D_k^{(t+1)} | X^{(t+1)}, M_k)\, \hat{\sigma}^{(\cdot t+1)}(X^{(t+1)})\, \hat{\sigma}^{(t)}(M_k) \prod_{i \neq k}^{K} \sum_{M_i} P(D_i^{(t+1)} | X^{(t+1)}, M_i)\, \hat{\sigma}^{(t)}(M_i)$$

   $(before\ normalization)$

2. Forward Inference memory requirements: $O((K+1)N^4)$. This is because there are N^2 states due to the square grid. The state transition model thus requires $O(N^4)$ space for each of the (K+1) variables.
   Unrolled DBN memory requirements: $O(T(K+1)N^4)$ using the naïve method. It requires $O(T(K+1)N^2)$ if caching the messages and evidence. The factor of T appears because there are now T time steps.

3. We can cache the calibrated messages and initialize the messages for the next cluster graph to these cached values instead of all 1's. This will result in faster convergence and thus faster calibration.

4. A scenario where inference on the unrolled DBN will produce more accurate localization results is in the case of sensor malfunction. This is because the unrolled DBN allows smoothing. Thus, prior beliefs can account for future observations. This is not possible in forward inference because there is no feedback from the current belief state to previous beliefs.

5. The Unrolled DBN results in the most accurate localization results. The Unrolled DBN has the sharpest peak in the final belief state. Sliding Window Inference takes the longest time. The first time step of Forward Inference with a single peak is Time 4.

6. For Forward Inference and Sliding Window Inference, the landmark beliefs change over time. For Unrolled DBN Inference the landmark beliefs are the same for each time slice. This is because Unrolled DBN Inference looks at the entire network at each time step,

and thus does not gain more information about static variables. On the other hand, FI and SW Inference gain more information at each step.

7. A window of size 5 gives the best localization results. A window of size 5 also produces the sharpest peak.
8. a) The scopes of cliques (D_j) in S need to create valid clique tree. Thus, each of the cliques need to satisfy the running intersection property, the family preservation property. To satisfy family preservation, for each $C_i^{(t)}$ there must exist a D_j such that D_j is a superset of $C_i^{(t)}$. Similarly, to extract $C_i^{(t+1)}$, for each $C_i^{(t+1)}$ there must exist a D_j such that D_j is a superset of $C_i^{(t+1)}$.

   b) Since $T^{(t+1)}$and $T^{(t)}$are alternate ways of representing $\hat{\sigma}^{(t+1)}\left(X^{(t+1)}, M\right)$ and $\hat{\sigma}^{(t)}\left(X^{(t)}, M\right)$, they can be calculated by extracting the require factors of $T^{(t+1)}$ from S, similar to how $\hat{\sigma}^{(t+1)}\left(X^{(t+1)}, M\right)$ would be extracted from S.

   i) $\pi_j = \prod_{C_i \subseteq \mathrm{D_j}} \frac{\beta_i^{(t)}}{\prod_{k=Pa\{i\}} \tau^{(t)}(S_{i,k})}$ where each β_i and $\tau(S_{i,j})$ are assigned to only one π_j in order to avoid double-counting.

   ii) Assuming $\mu(S_{i,j}^{'})$ represent the final calibrated messages, then the calibrated potentials for each $C_i^{(t+1)}$ can be extracted using:

   $$\beta_i^{(t+1)} = \sum_{D_j - C_i} \pi_j \prod_{k \in Nb\{D_j\}} \mu(S_{k,j}^{'})$$

   $$\tau^{(t+1)}(S_{i,j}) = \sum_{C_i - S_{i,j}} \beta_i^{(t+1)}$$

   c) The inference developed in parts (a) and (b) will not work here in general. This is because the interactions between variables that the tree structure $T^{(t)}$ represents may change over time. Thus, the clique tree may no longer satisfy the family preserving property specified in part (a).