Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Adaptively stacking ensembles for influenza forecasting: Supplementary Information

Thomas McAndrew*[1,2] | Nicholas G. Reich[2]

[1]Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts Amherst, Amherst, Massachusetts,, United States

[2]College of Health, Lehigh University, Bethlehem, Pennsylvania, United States

Correspondence
*Thomas McAndrew, Lehigh University Bethlehem, Pennsylvania, United States of America. Email: mcandrew@lehigh.edu

Summary

Seasonal influenza infects between 10 and 50 million people in the United States every year. Accurate forecasts of influenza and influenza-like illness (ILI) have been named by the CDC as an important tool to fight the damaging effects of these epidemics. Multi-model ensembles make accurate forecasts of seasonal influenza, but current operational ensemble forecasts are static: they require an abundance of past ILI data and assign fixed weights to component models at the beginning of a season, but do not update weights as new data on component model performance is collected. We propose an adaptive ensemble that (i) does not initially need data to combine forecasts and (ii) finds optimal weights which are updated week-by-week throughout the influenza season. We take a regularized likelihood approach and investigate this regularizer's ability to impact adaptive ensemble performance. After finding an optimal regularization value, we compare our adaptive ensemble to an equal-weighted and static ensemble. Applied to forecasts of short-term ILI incidence at the regional and national level, our adaptive model outperforms an equal-weighted ensemble and has similar performance to the static ensemble using only a fraction of the data available to the static ensemble. Needing no data at the beginning of an epidemic, an adaptive ensemble can quickly train and forecast an outbreak, providing a practical tool to public health officials looking for a forecast to conform to unique features of a specific season.

KEYWORDS:
Combination forecasting; Forecast aggregation; Influenza; Statistics; Public health

1 | COMPUTING $Q(\pi)$ AND $Q(Z)$ FOR THE DEGENERATE VARIATIONAL MIXTURE MODEL (DEVI-MM)

The functional forms for $q(Z)$ and $q(\pi)$ are computed. Readers interested in more details, and theoretical background, should consult [1,2,3,4]. In particular,[2] gives a brief introduction to variational inference focused on the applied statistician, while [3,4] provide more theoretical details. The most detailed material can be found in [1].

We find the $q(Z)$ and $q(\pi)$ that maximize the lower bound $\mathcal{L}(q)$ by taking advantage of our factored q and using iterated expectations.

$$\mathcal{L}(q) = E_{\pi,Z}\left\{\log\left[p(\mathcal{D}, Z, \pi)\right] - \log[q(\pi)] - \log[q(Z)]\right\} \tag{1}$$

Maximizing $q(\pi)$, we can take the iterated expectation

$$\begin{aligned}\max_{q(\pi)} \mathcal{L}(q) &= \max_{q(\pi)} E_{\pi|Z} E_Z \left\{\log\left[p(\mathcal{D}, Z, \pi)\right] - \log[q(\pi)] - \log[q(Z)]\right\} \\ &= \max_{q(\pi)} E_{\pi|Z} \left\{E_Z \log\left[p(\mathcal{D}, Z, \pi)\right] - \log[q(\pi)]\right\} \\ &= \min_{q(\pi)} \mathrm{KL}\left\{\log q(\pi) || E_Z \log\left[p(\mathcal{D}, Z, \pi)\right]\right\}\end{aligned}$$

The Kullback-Leibler divergence, taking values from 0 to ∞, is minimized when

$$\log q(\pi) = E_Z \log\left[p(\mathcal{D}, Z, \pi)\right]$$

or when

$$q(\pi) \propto \exp\left\{E_Z \log\left[p(\mathcal{D}, Z, \pi)\right]\right\}.$$

Maximizing Z follows a similar pattern. The optimal hidden distributions of $q(\pi)$ and $q(Z)$ can then be computed

$$\begin{aligned}q(\pi) &\propto \exp\left\{E_Z \log\left[p(D, Z, \pi)\right]\right\} \\ q(Z) &\propto \exp\left\{E_\pi \log\left[p(D, Z, \pi)\right]\right\}.\end{aligned}$$

We first compute $q(\pi)$, expanding the complete loglikelihood, taking the expectation over Z, and recognizing this expectation as a specific distribution. Here we don't explicitly describe π's dependence on t for convenience.

$$\begin{aligned}\log q(\pi) &\propto \mathbb{E}_Z \log\left[p(D, Z, \pi)\right] \\ &= \sum_{t=1}^{T}\sum_{m=1}^{M} \mathbb{E}\left[z(m,t)\right] \log\left[\pi_m f_m(y_t)\right] + \sum_{m=1}^{M} \left[\alpha(t) - 1\right] \log(\pi_m) - \log\left\{\eta\left[\alpha(t)\right]\right\} \\ &= \sum_{t=1}^{T}\sum_{m=1}^{M} \mathbb{E}\left[z(m,t)\right] \log\left(\pi_m\right) + \sum_{m=1}^{M} \left[\alpha(t) - 1\right] \log\left(\pi_m\right) \\ &= \sum_{m=1}^{M} \log\left(\pi_m\right) \left\{\alpha(t) + \sum_{t=1}^{T} \mathbb{E}\left[z(m,t)\right] - 1\right\} \\ &= \sum_{m=1}^{M} \log\left(\pi_m\right) \left\{\alpha(t) + \sum_{t=1}^{T} r(m,t) - 1\right\},\end{aligned}$$

where η is the normalizing constant for the Dirichlet distribution and $r(m,t)$ the expected value of the indicator variable $z_{m,t}$, the probability model m generated the ILI value at time t. Studying the form of $\log q(\pi)$,we recognize π is Dirichlet distributed

$$\begin{aligned}q(\pi) &\sim \mathrm{Dir}\left(\gamma\right) \\ \gamma\left[m\right] &= \alpha(t) + \sum_{t=1}^{T} r(m,t)\end{aligned}$$

The same procedure can also be applied to compute $q(Z)$:

$$\begin{aligned}\log q(Z) &\propto \mathbb{E}_\pi \log\left[p(D, Z, \pi)\right] \\ &= \sum_{t=1}^{T}\sum_{m=1}^{M} z(m,t) \left\{\mathbb{E}_\pi \log\left(\pi_m\right) + \log\left[f_m(y_t)\right]\right\} + \\ &\mathbb{E}_\pi \sum_{m=1}^{M} \left[\alpha(t) - 1\right] \log(\pi_m) - \log\left[\eta(\alpha)\right] \\ &\propto \sum_{t=1}^{T}\sum_{m=1}^{M} z(m,t) \left\{\mathbb{E}_\pi \log\left[\pi_m\right] + \log\left[f_m(y_t)\right]\right\} \\ q(Z_{t,m}) &\propto \exp\left\{\mathbb{E}_\pi \log\left(\pi_m\right) + \log\left[f_m(y_t)\right]\right\},\end{aligned} \tag{2}$$

and we recognize $q(Z_{t,m})$ is Bernoulli distributed, and with the additional constraint that all indicators must sum to one for every time period (t), we see Z_t is multinomial for every t.

$$q(Z_{t,m}) \sim \text{Bern}\left[r(m,t)\right] \tag{3}$$

$$r(m,t) = \frac{\exp\left\{\mathbb{E}_\pi \log\left(\pi_m\right) + \log\left[f_m(y_t)\right]\right\}}{\sum_{m=1}^{M} \exp\left\{\mathbb{E}_\pi \log\left(\pi_m\right) + \log\left[f_m(y_t)\right]\right\}}. \tag{4}$$

Although burdensome upfront, this approximate procedure drastically reduces computational time, compared to more intense monte carlo sampling techniques, and gives a good approximate answer, only assuming Z and π independent from one another. Also note this mixture model algorithm (both EM and VI), unlike a typical Gaussian mixture model, cannot manipulate the paramters control the component model distributions. But this inability to access the component model parameters opens up greater opportunities for other forecasters to submit models, requiring every forecast model to supply just 1, 2 ,3 , and 4 week ahead forecasts.

2 | COMPUTING $E[\log(\pi)]$

Variational inference over hidden variables (Z, π) requires computing the expected value of the log of π's, a natural consequence of adapting the Dirichlet distribution to exponential form. Exponential form rewrites a probability distribution in the form

$$p\left[\pi|\eta(\alpha)\right] = h(\pi)e^{\sum_k \eta_k(\alpha)T_k(\pi) - c(\eta)}. \tag{5}$$

where $c(\pi)$ is a normalizing constant. Two facts about exponential form lead to an analytic formula for $E(\log \pi)$: taking the gradient of $c(\pi)$ and relating the set of sufficient statistics $T(\pi)$ with this expectation. Starting with the first fact. If $c(\pi)$ is a normalizing constant, then it must equal

$$c(\eta) = \log\left(\int h(\pi)e^{\eta'(\alpha)T(\pi)}\, d\pi\right), \tag{6}$$

and it's gradient must equal

$$\nabla_\eta\left[c(\eta)\right] = \int T_k(\pi)\frac{h(\pi)e^{\eta'(\alpha)T(\pi)}}{\int h(\pi)e^{\eta'(\alpha)T(\pi)}}\, d\pi \tag{7}$$

$$= \int T(\pi)p(\pi) = E[T(\pi)]. \tag{8}$$

A powerful consequence of exponential form, the gradient of the normalizing constant equals the expected value of the distribution's sufficient statistics. The Second fact. Working with the log likelihood of a distribution in exponential form,

$$\log\left[\prod_{n=1}^{N} p(\pi|\alpha)\right] = N\log h(\pi) + N\sum_k \eta_k(\alpha_k)T_k(\pi) - Nc(\pi), \tag{9}$$

taking the gradient and setting equal to 0,

$$\nabla_\eta \log\left[\prod_{n=1}^{N} p(\pi|\alpha)\right] = \sum T(\pi_n) - N\nabla c(\eta) = 0 \tag{10}$$

$$\nabla c(\eta) = E\left[T(\eta)\right] = \frac{1}{N}T(\pi). \tag{11}$$

The expected value of the distribution's sufficient statistics are equal to the gradient of $c(\eta)$. If the Dirichlet's sufficient statistics take the form $\log(\pi)$, then we only need to take the gradient of the normalizing constant to find an analytic expression.

Looking at the Dirichlet distribution, the loglikelihood equals

$$\log\left[\prod p(\pi_n|\alpha)\right] = N\log\Gamma\left(\sum_\alpha \alpha_k\right) - N\sum_k \log\Gamma\left(\alpha_k\right) + N\sum_k(\alpha_k - 1)\log(\pi_k) \tag{12}$$

the sufficient statistics for π_k take the form $\log\left(\pi_k\right)$. Taking the gradient then will provide an analytic expression for computing the $\log \pi$'s expected value.

$$\nabla_\eta \log\left[\prod p(\pi_n|\alpha)\right] = N\psi\left(\sum_\alpha \alpha_k\right) - N\sum_k \psi\left(\alpha_k\right) + N\log(\pi_k), \tag{13}$$

where ψ is the digamma function. Then the expected value of $\log(\pi)$ equals a difference of digamma functions

$$E\left[\pi_k\right] = \psi\left(\alpha_k\right) - \psi\left(\sum_\alpha \alpha_k\right). \tag{14}$$

This formula can be used to compute the responsibility function (see Algorithm 2 step 7).

3 | CONVEXITY ANALYSIS

The problem of finding an optimal set of mixture weights can be recast as a constrained optimization problem. Attempting to optimize a convex function over a convex set will prove a global optima exists, and showing strict convexity will prove the a unique vector obtains this optimum.

The original problem searches for weights that maximize a loglikelihood whose sum is constrained to equal one,

$$\begin{aligned} &\max \sum_{t=1}^{T} \log\left[\sum_{m=1}^{M} \pi_m f_m(y_t)\right] \\ &\text{subject to} \\ &\sum_{m=1}^{M} \pi_m = 1, \end{aligned} \tag{15}$$

and can be converted to a Langragian with a single constraint (λ)

$$\mathcal{L}(\pi, \lambda) = \sum_{t=1}^{T} \log\left[\sum_{m=1}^{M} \pi_m f_m(y_t)\right] + \lambda\left(\sum_{m=1}^{M} \pi_m - 1\right).$$

Knowing the solution π needs to lie in the M dimensional simplex, a convex set, we only need to show the above function (15) is convex. Given (15) is differentiable at least twice, we will appeal to a second-order condition for convexity—proving the Hessian is positive semidefinite. After proving the Hessian is positive semidefinite, going further and showing the Hessian is in fact positive definite will prove a unique vector π is a the global optimum of our objective function (15).

First, we compute the m^{th} element of the gradient for (15),

$$\nabla_{\pi_m} f(\pi) = \sum_{t=1}^{T} \frac{f_m(i_i)}{\sum_{m=1}^{M} \pi_m f_m(y_t)}.$$

Then the (m, n) entry of the Hessian is

$$H(m,n) = -\sum_{t=1}^{T} \frac{f_m(y_t) f_n(y_t)}{\left[\sum_{m=1}^{M} \pi_m f_m(y_t)\right]^2}.$$

The Hessian is always negative, and if we minimized the negative loglikelihood (instead of maximizing the positive logelikihood) would see the Hessian is a positive semidefinite matrix. But we can prove more by rewriting H

$$H = F'F,$$

where the $[m, t]$ element

$$F[m,t] = \frac{f_m(y_t)}{\sum_{m=1}^{M} \pi_m f_m(y_t)}.$$

Positive definite matrices can always be written as a transposed matrix times the matrix itself, and so H must also be positive definite. Our convex optimization problem must then have a global optimum, attained by a unique vector π. The inability to change forecasting models is a limitation, but allows us to guarantee a global optima. When the component model parameters can also be updated, no such guarantee exists.

4 | SIMPLEX PLOTS OF ALL SEASONS

Regularizing the adaptive ensemble suppresses large swings in weight assignment, independent of season. The **adaptive**$_{\textbf{over}}$ ensemble stays closer to equal weighting throughout the season compared to the **adaptive**$_{\textbf{opt}}$ and **adaptive**$_{\textbf{non}}$ ensembles. If one exists, there is no strong relationship between the adaptive and static ensemble weighting. Some seasons (2014/2015 and 2016/2017) show higher variability in weight assignment than others. The optimal weight assignments given data from all previous seasons (Static MLE) does not carry forward to the optimal weight assignment for the adaptive ensemble at the end of the current season.

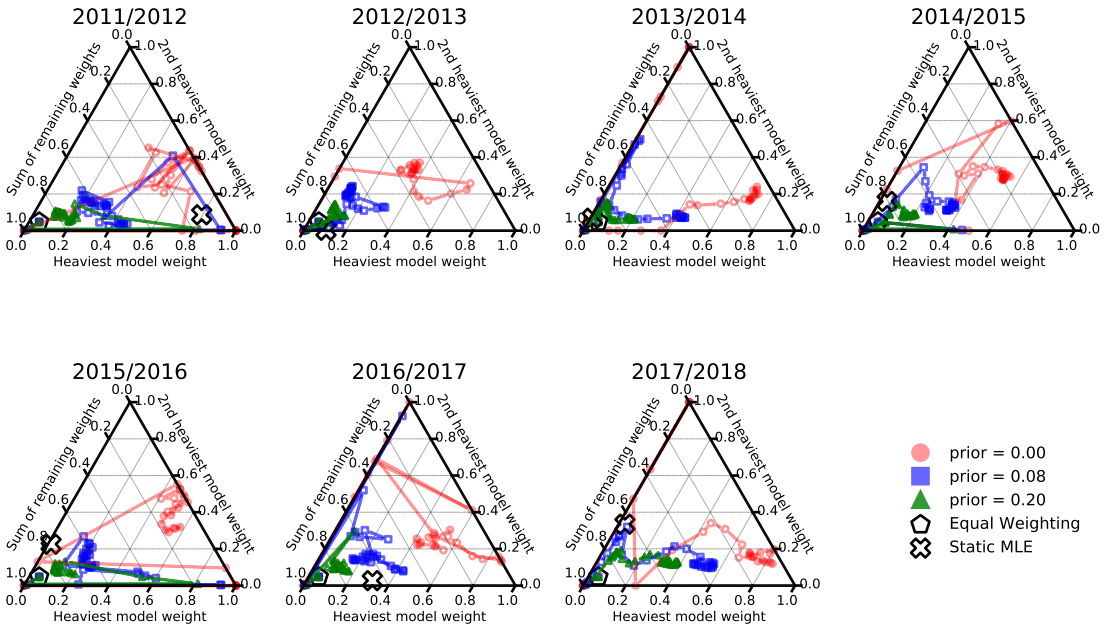

FIGURE 1 Component model weights for an adaptive ensemble with 0% (non), 8% (opt), and 20% (over) priors plotted over epidemic week and stratified by season. Equal weighting is represented by a pentagon and the Static ensemble weights represented by an X. All ensembles start at an equal weighted triple of the 1st and 2nd highest weighted component model, and the sum of the remaining components at the end the season ($\pi_t^{(1)}, \pi_t^2, \sum_{j=3}^{M} \pi_t^{(j)}$), and move from week to week as new data is received and the adaptive ensemble re-estimates component model weight assignments.

5 | TABULATED COMPARISONS OF ADAPTIVE, STATIC, AND EQUAL LOG SCORES

The **adaptive$_{opt}$** ensemble consistently outperforms the EW ensemble, and performs similarly to the static ensemble. When compared to the EW model, the adaptive cannot show statistical significance in the 2012/2013 season, but all differences are in the positive direction, favoring the adaptive model. The adaptive and static comparisons are more even. Some season, regions, and targets favor the static ensemble, others the adaptive model, but absolute differences are small. Despite markedly less data, the adaptive ensemble bests the EW ensemble and shows similar performance to the static.

	Adaptive$_{opt}$ - EW			**Adaptive$_{opt}$ - Static**		
	β (95CI)	p	$p_{permutation}$	β (95CI)	p	$p_{permutation}$
Season						
2011/2012	0.13 (0.08, 0.19)	<0.01	< 0.01	0.02 (-0.04, 0.08)	0.51	0.58
2012/2013	0.06 (0.00, 0.12)	0.04	0.21	0.02 (-0.04, 0.08)	0.48	0.52
2013/2014	0.10 (0.04, 0.15)	<0.01	0.04	0.00 (-0.06, 0.06)	0.96	0.97
2014/2015	0.14 (0.09, 0.20)	<0.01	< 0.01	0.03 (-0.03, 0.09)	0.30	0.36
2015/2016	0.13 (0.07, 0.18)	<0.01	0.01	-0.02 (-0.08, 0.04)	0.54	1.00
2016/2017	0.11 (0.06, 0.17)	<0.01	0.01	-0.04 (-0.10, 0.02)	0.20	1.00
2017/2018	0.21 (0.15, 0.26)	<0.01	0.00	-0.01 (-0.06, 0.05)	0.86	1.00
Region						
HHS1	0.16 (0.11, 0.22)	<0.01	< 0.01	0.01 (-0.05, 0.07)	0.80	0.83
HHS2	0.11 (0.06, 0.17)	<0.01	< 0.01	-0.01 (-0.07, 0.06)	0.85	1.00
HHS3	0.13 (0.08, 0.18)	<0.01	< 0.01	0.04 (-0.03, 0.10)	0.28	0.41
HHS4	0.12 (0.07, 0.17)	<0.01	< 0.01	-0.01 (-0.08, 0.05)	0.68	1.00
HHS5	0.13 (0.07, 0.18)	<0.01	< 0.01	-0.05 (-0.11, 0.01)	0.12	1.00
HHS6	0.12 (0.07, 0.18)	<0.01	0.001	-0.04 (-0.11, 0.02)	0.17	1.00
HHS7	0.13 (0.08, 0.18)	<0.01	< 0.01	0.05 (-0.01, 0.12)	0.09	0.20
HHS8	0.14 (0.09, 0.20)	<0.01	< 0.01	0.00 (-0.06, 0.06)	0.99	1.00
HHS9	0.08 (0.02, 0.13)	<0.01	0.01	0.02 (-0.04, 0.08)	0.55	0.60
HHS10	0.12 (0.07, 0.17)	<0.01	< 0.01	-0.04 (-0.11, 0.02)	0.19	1.00
Nat	0.13 (0.07, 0.18)	<0.01	< 0.01	0.06 (-0.01, 0.12)	0.08	0.17
Target						
1 week ahead	0.16 (0.11, 0.22)	<0.01	< 0.01	0.06 (0.00, 0.11)	0.07	0.18
2 week ahead	0.13 (0.08, 0.19)	<0.01	< 0.01	-0.01 (-0.07, 0.05)	0.81	1.00
3 week ahead	0.11 (0.06, 0.17)	<0.01	< 0.01	-0.01 (-0.07, 0.05)	0.67	1.00
4 week ahead	0.09 (0.03, 0.14)	<0.01	0.02	-0.03 (-0.09, 0.03)	0.35	1.00

TABLE 1 Random effects regressions compared log scores between the **adaptive$_{opt}$** vs equally-weighted and **adaptive$_{opt}$** vs static ensembles. The model included an intercept, and separate random effect for: season, region, and target. The dependent variables is the difference in log scores paired by season-region-target-epidemic week. Conditional mean, 95%CI, asymptotic, and a permutation based p-value are reported.

	Adaptive$_{non}$ - EW			**Adaptive$_{non}$ - Static**		
	β (95CI)	p	$p_{permutation}$	β (95CI)	p	$p_{permutation}$
Season						
2011/2012	0.09 (0.00, 0.18)	0.04	0.16	-0.02 (-0.11, 0.06)	0.63	1.00
2012/2013	0.00 (-0.09, 0.09)	0.95	1.00	-0.05 (-0.13, 0.04)	0.28	1.00
2013/2014	0.04 (-0.13, 0.05)	0.35	1.00	-0.12 (-0.21, -0.04)	<0.01	1.00
2014/2015	0.12 (0.03, 0.21)	0.01	0.06	0.01 (-0.07, 0.10)	0.79	0.82
2015/2016	0.06 (-0.03, 0.15)	0.20	0.37	-0.09 (-0.17, 0.00)	0.04	1.00
2016/2017	0.09 (0.00, 0.17)	0.06	0.18	-0.06 (-0.14, 0.03)	0.18	1.00
2017/2018	0.10 (0.01, 0.19)	0.03	0.14	-0.10 (-0.18, -0.01)	0.02	1.00
Region						
HHS1	0.15 (0.05, 0.24)	<0.01	0.02	-0.01 (-0.10, 0.08)	0.89	1.00
HHS2	0.06 (-0.03, 0.16)	0.18	0.32	-0.06 (-0.15, 0.04)	0.24	1.00
HHS3	0.10 (0.00, 0.19)	0.04	0.14	0.00 (-0.09, 0.09)	1.00	1.00
HHS4	0.09 (0.00, 0.18)	0.06	0.15	-0.04 (-0.13, 0.05)	0.36	1.00
HHS5	0.03 (-0.07, 0.12)	0.57	0.70	-0.15 (-0.24, -0.06)	<0.01	1.00
HHS6	0.00 (-0.09, 0.09)	0.99	1.00	-0.13 (-0.22, -0.04)	<0.01	1.00
HHS7	0.07 (-0.02, 0.17)	0.13	0.27	0.00 (-0.10, 0.09)	0.92	1.00
HHS8	0.11 (0.01, 0.20)	0.02	0.11	-0.03 (-0.13, 0.06)	0.46	1.00
HHS9	0.07 (-0.16, 0.03)	0.16	1.00	-0.13 (-0.22, -0.04)	0.01	1.00
HHS10	0.05 (-0.04, 0.14)	0.30	0.42	-0.11 (-0.20, -0.02)	0.02	1.00
Nat	0.06 (-0.03, 0.16)	0.18	0.30	-0.01 (-0.10, 0.08)	0.89	1.00
Target						
1 week ahead	0.09 (0.01, 0.16)	0.03	0.04	-0.02 (-0.09, 0.06)	0.70	1.00
2 week ahead	0.06 (-0.02, 0.14)	0.13	0.15	-0.07 (-0.15, 0.00)	0.06	1.00
3 week ahead	0.06 (-0.02, 0.14)	0.14	0.15	-0.06 (-0.14, 0.01)	0.11	1.00
4 week ahead	0.03 (-0.05, 0.11)	0.42	0.44	-0.09 (-0.17, -0.01)	0.03	1.00

TABLE 2 Random effects regressions compared log scores between the **adaptive$_{non}$** vs equally-weighted and **adaptive$_{non}$** vs static ensembles. The model included an intercept, and separate random effect for: season, region, and target. The dependent variables is the difference in log scores paired by season-region-target-epidemic week. Conditional mean, 95%CI, asymptotic, and a permutation based p-value are reported.

	Adaptive$_{over}$ - EW			**Adaptive$_{over}$ - Static**		
	β (95CI)	p	$p_{permutation}$	β (95CI)	p	$p_{permutation}$
Season						
2011/2012	0.11 (0.06, 0.16)	<0.01	0.01	0.00 (-0.05, 0.06)	0.94	0.95
2012/2013	0.06 (0.01, 0.10)	0.03	0.23	0.02 (-0.04, 0.07)	0.54	0.62
2013/2014	0.07 (0.02, 0.12)	0.01	0.12	0.01 (-0.07, 0.04)	0.62	1.00
2014/2015	0.12 (0.08, 0.17)	<0.01	0.01	0.01 (-0.04, 0.07)	0.62	0.72
2015/2016	0.11 (0.06, 0.16)	<0.01	0.02	0.04 (-0.09, 0.02)	0.18	1.00
2016/2017	0.07 (0.03, 0.12)	<0.01	0.12	0.06 (-0.12, -0.01)	0.02	1.00
2017/2018	0.18 (0.14, 0.23)	<0.01	0.00	0.03 (-0.08, 0.03)	0.35	1.00
Region						
HHS1	0.13 (0.09, 0.18)	<0.01	0.00	-0.02 (-0.08, 0.04)	0.45	1.00
HHS2	0.10 (0.05, 0.15)	<0.01	0.00	-0.02 (-0.08, 0.04)	0.57	1.00
HHS3	0.11 (0.06, 0.15)	<0.01	0.00	0.01 (-0.05, 0.07)	0.69	0.78
HHS4	0.10 (0.06, 0.15)	<0.01	0.00	-0.03 (-0.09, 0.03)	0.29	1.00
HHS5	0.11 (0.07, 0.16)	<0.01	0.00	-0.06 (-0.12, 0.00)	0.04	1.00
HHS6	0.07 (0.03, 0.12)	<0.01	0.00	-0.06 (-0.12, 0.00)	0.06	1.00
HHS7	0.11 (0.07, 0.16)	<0.01	0.00	0.04 (-0.02, 0.10)	0.22	0.40
HHS8	0.12 (0.07, 0.16)	<0.01	0.00	-0.03 (-0.09, 0.03)	0.38	1.00
HHS9	0.08 (0.03, 0.12)	<0.01	0.01	0.02 (-0.04, 0.08)	0.54	0.66
HHS10	0.10 (0.06, 0.15)	<0.01	0.00	-0.06 (-0.12, 0.00)	0.05	1.00
Nat	0.11 (0.06, 0.15)	<0.01	0.00	0.04 (-0.02, 0.10)	0.22	0.36
Target						
1 week ahead	0.13 (0.08, 0.18)	<0.01	0.00	0.03 (-0.03, 0.08)	0.33	0.43
2 week ahead	0.11 (0.06, 0.16)	<0.01	0.00	-0.03 (-0.08, 0.03)	0.35	1.00
3 week ahead	0.10 (0.05, 0.14)	<0.01	0.00	-0.03 (-0.08, 0.03)	0.33	1.00
4 week ahead	0.08 (0.03, 0.12)	<0.01	0.01	-0.04 (-0.09, 0.02)	0.19	1.00

TABLE 3 Random effects regressions compared log scores between the **adaptive$_{over}$** vs equally-weighted and **adaptive$_{over}$** vs static ensembles. The model included an intercept, and separate random effect for: season, region, and target. The dependent variables is the difference in log scores paired by season-region-target-epidemic week. Conditional mean, 95%CI, asymptotic, and a permutation based p-value are reported.

6 | REGULARIZATION IMPROVES STATIC ENSEMBLE

Regularizing ensemble weights improves both adaptive and static ensemble performance (Fig. 2). The static ensemble achieves peak performance with a smaller prior than the adaptive. This smaller prior reflects that the static model is trained on data from past seasons of finalized ILI data. The adaptive model, finding peak performance for a larger prior, needs to account for the lack of training data and the revision-prone state of the data mid-season.

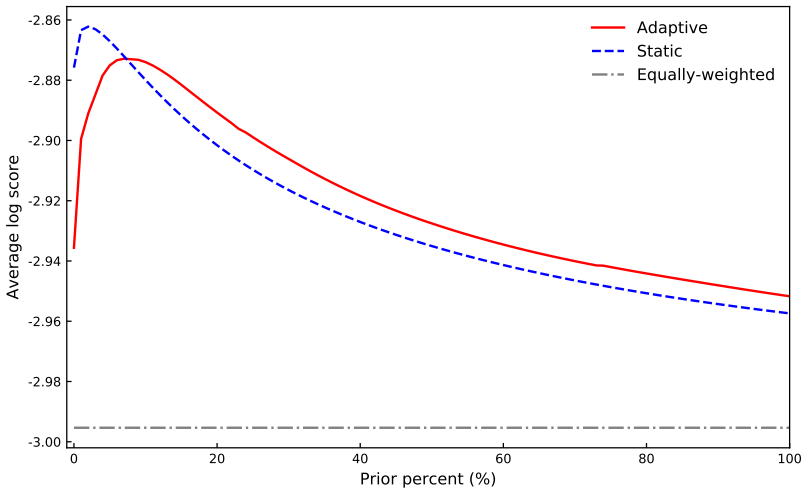

FIGURE 2 The average log score for the equally-weighted ensemble, and adaptive, static ensembles for prior from 0% to 100% by 1%. The log score is averaged over season, region, and target. The highest average log score corresponds to a larger prior than the static ensemble. But both the static and adaptive ensemble benefit from regularization.

7 | SUPPLEMENTARY FIGURE

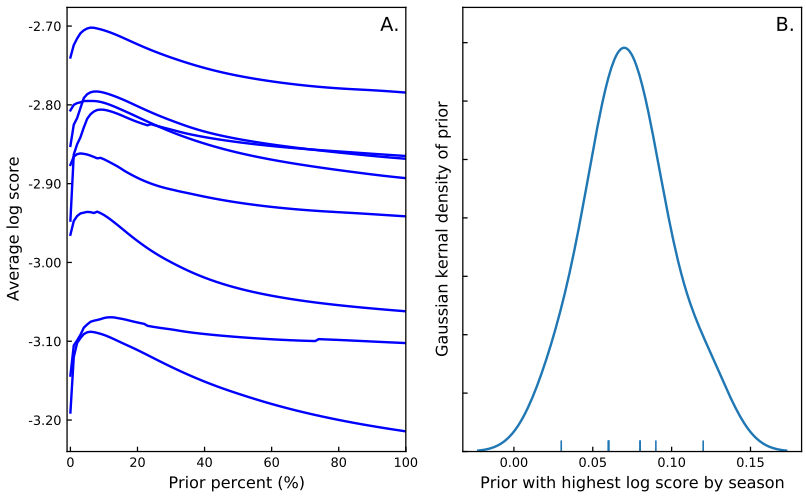

FIGURE 3 (A.) The adaptive ensemble is fit for prior percentages from 0% to 100% by 1% and the average log score is computed, and stratified by seasons 2010/2011-2017/2018. (B.) the distribution of priors corresponding to the highest log score per season. (A.) Each seasons shows a similar trend in adaptive ensemble log score using different prior percentages. Priors close to 0% produce small log scores, a peak log score occurs near a prior of 8%, and then log scores decrease with larger priors. (B.) The 25th and 75th percentile of priors corresponding to peak log scores are 6% and 8.25%. The smallest prior equals 3% and largest equals 12%. The large probability near a prior percentage of 8%, across influenza seasons, suggests an optimal weighting may lie near 8%, was not specific to the 2010/2011 season analysis, and could work well as a prior for future seasons.

References

1. Rustagi JS. Variational methods in statistics. In: Academic Press. 1976.

2. Blei DM, Kucukelbir A, McAuliffe JD. Variational inference: A review for statisticians. *Journal of the American Statistical Association* 2017; 112(518): 859–877.

3. Bishop CM. *Pattern recognition and machine learning*. springer . 2006.

4. Murphy K. Machine learning: a probabilistic approach. *Massachusetts Institute of Technology* 2012: 1–21.

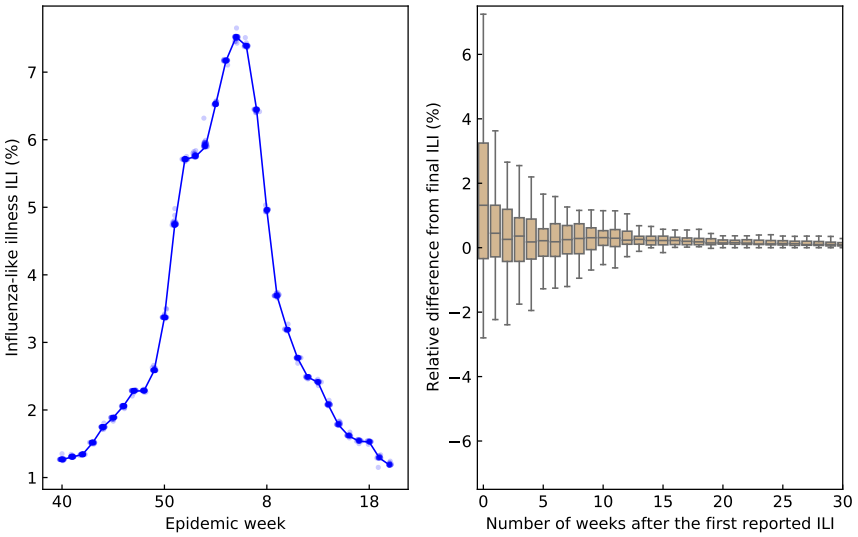

FIGURE 4 (A.) The final influenza-like illness (blue line) and all revised influenza-like illness values (blue dots) by epidemic week for the 2017/2018 season. (B.) The relative difference between the final influenza-like illness and revised influenza-like illness by the number of weeks after the first reported influenza-like illness value. Revised values show the largest differences close to the peak final influenza-like illness. The relative difference is highest in the first few weeks after an influenza-like illness value is reported. After 10 weeks the revised influenza-like illness is likely within 1% of the final value. This is one factor that may contribute to difficulty assessing true component model performance throughout the influnza season.