<div align="center">

Web Material

Type- and Subtype-Specific Influenza Forecast

Sasikiran Kandula, Wan Yang, Jeffrey Shaman

</div>

Web Appendix 1

Filters

Generally, Kalman filters assume normality of both the likelihood and prior distributions during an update. Differences among the ensemble filter algorithms manifest in the means by which the update is specified. The ensemble Kalman filter is a stochastic, perturbed observation form of the Kalman filter in which the update of each ensemble member is computed using the current observation plus Gaussian random noise (1). That is, the posterior for each ensemble member is simply the weighted average of the prior for that ensemble member and the observation plus random noise with variance equal to the observational error variance. The weights themselves are calculated as ratios of the ensemble prior variance and the observational error variance. The ensemble adjustment Kalman filter employs a deterministic algorithm to compute the ensemble posterior mean and variance (2). At each update, the ensemble adjustment Kalman filter algorithm aligns the first two ensemble posterior moments with those predicted by Bayes theorem.

Unlike the ensemble Kalman filter and ensemble adjustment Kalman filter, the rank histogram filter does not impose a Gaussian structure on the prior, observations and posterior (3); rather, this filter employs an algorithm that creates an approximate probability distribution by ordering (i.e. ranking) the ensemble prior. In this fashion, the rank histogram filter admits non-Gaussian distributions, thus relaxing the normality assumption inherent to most Kalman filters.

Multiplicative inflation (2, 4) was applied following the assimilation of each weekly observation of ILI+, H1N1+, H3N2+ or B+. The inflation was used to counter the ensemble filter tendency toward 'filter divergence', which occurs when the prior ensemble spread becomes spuriously small. In the absence of inflation, the system may give too little weight to the observations and thus diverge from the true trajectory.

Scaling

We map observations to incidence through a scaling factor, γ (5). Specifically, the observed metric, ILI+, H1N1+, H3N2+ or B+, is a measure of patients presenting with influenza (or a specific type/subtype) per 100,000 patient visits. This fraction neglects those not seeking medical attention and is not scaled to the whole population. Mapping from ILI+, H1N1+, H3N2+ or B+ to simulated influenza is formalized using Bayes theorem. For example, ILI+ is simply an estimate of the probability for a given week that a person seeking medical treatment, m, has influenza, i.e. $p(i \mid m)$. By Bayes theorem, ILI+ is then

$$\text{ILI+} \gg p(i \mid m) = \frac{p(i)\, p(m \mid i)}{p(m)} \qquad [1]$$

<div align="center">

1

</div>

where $p(i)$ is the probability of getting influenza in a given week (i.e. influenza incidence rate), $p(m|i)$ is the probability of seeking medical attention given infection with influenza, and $p(m)$ is the probability that anyone seeks medical attention for any reason. Equation 1 can be rearranged:

$$p(i) = \frac{p(m)}{p(m|i)} p(i|m) \approx g\text{ILI+} \qquad [2]$$

where $g = p(m)/p(m|i)$. That is, the probability of incident influenza infection in the general population, $p(i)$, as simulated in the forecast model, is approximately equal to ILI+ scaled by the probability that a person seeks medical attention for any reason, $p(m)$, and the probability that a person with influenza seeks medical attention, $p(m|i)$. Similar arguments apply for H1N1+, H3N2+ and B+. For all forecasts scaling was fixed at $\gamma = 2.5$.

References

1. Evensen G. Data Assimilation: The ensemble Kalman filter. 2nd ed. New York: Springer. 2009.
2. Anderson JL. An Ensemble Adjustment Kalman Filter for Data Assimilation. *Mon. Weather Rev.* 2001;129(12):2884-2903.
3. Anderson JL. A non-Gaussian ensemble filter update for data assimilation. *Mon Weather Rev.* 2010;138(11): 4186-4198.
4. Shaman J, Karspeck A. Forecasting seasonal outbreaks of influenza. *Proc Natl Acad Sci USA.* 2012;109(50):20425-20430.
5. Shaman J, Karspeck A, Yang W, et al. Real-time influenza forecasts during the 2012-2013 season. *Nat Commun.* 2013; 4:2837.
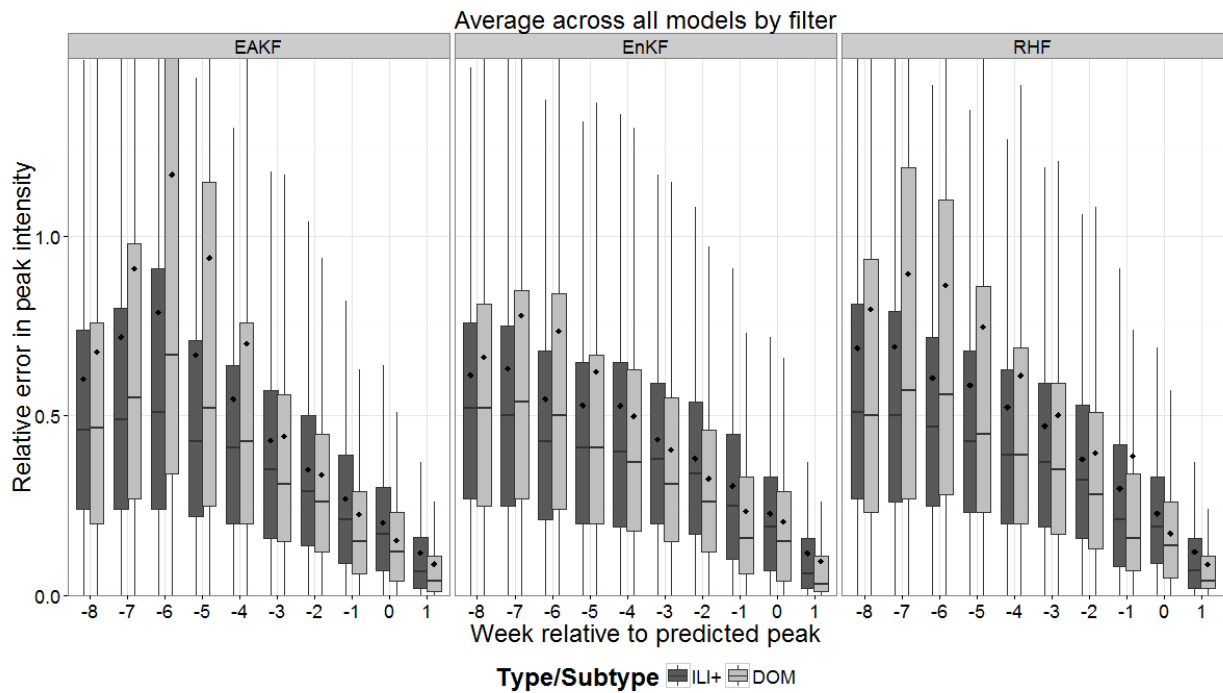
Web Figure 1. As for Figure 1, but stratified by filter type. Box and whisker plots show the median (thick horizontal line), 25th and 75th percentiles (the bottom and top extent of the box) and extrema (the whiskers) of the absolute error of peak week timing predictions. The black dots indicate the mean absolute error for each box and whisker.
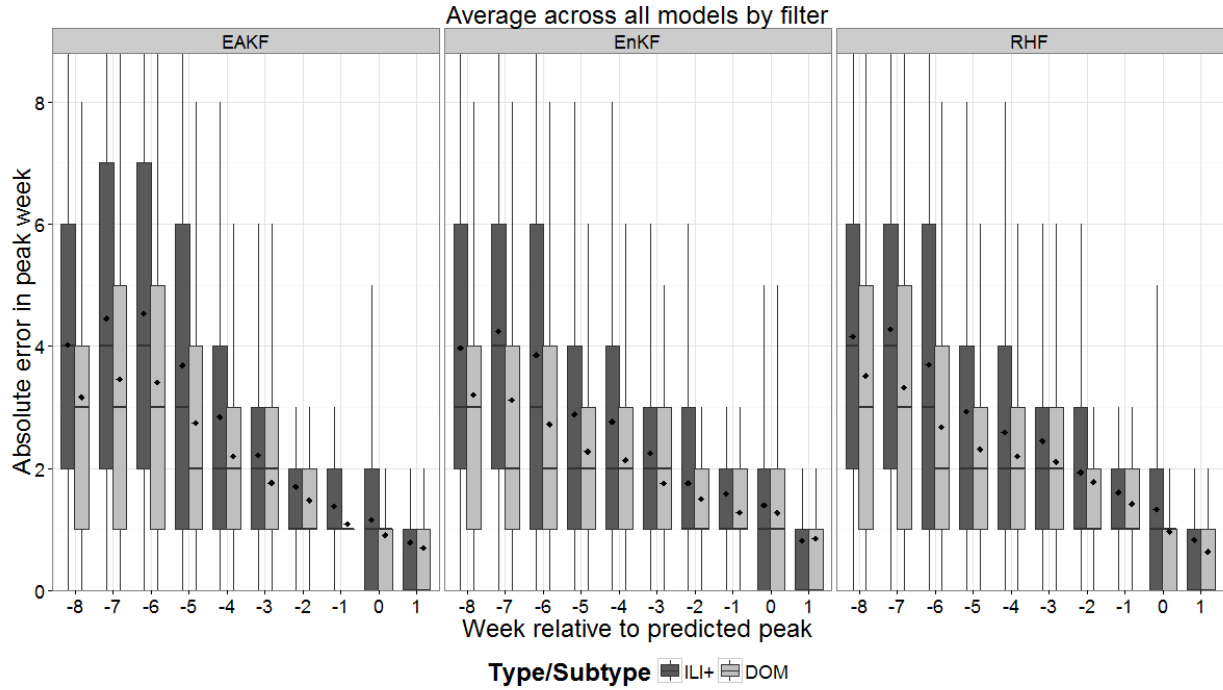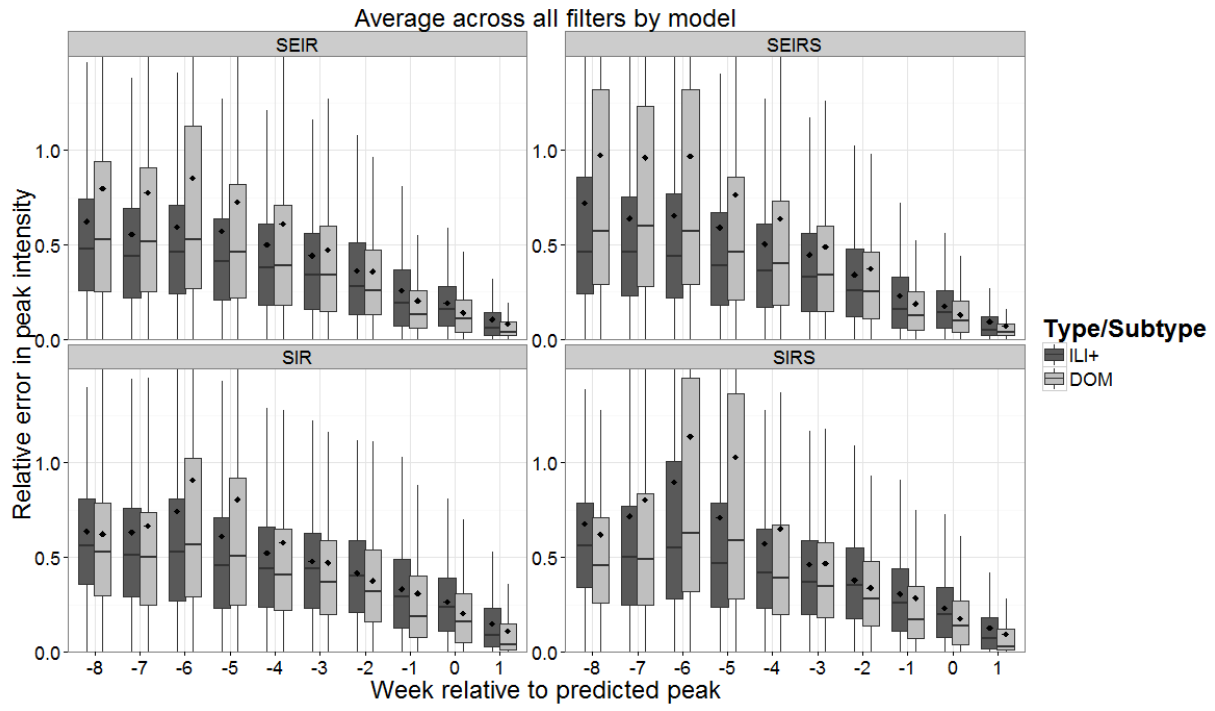
**Web Figure 2.** As for Figure 1, but stratified by model type.
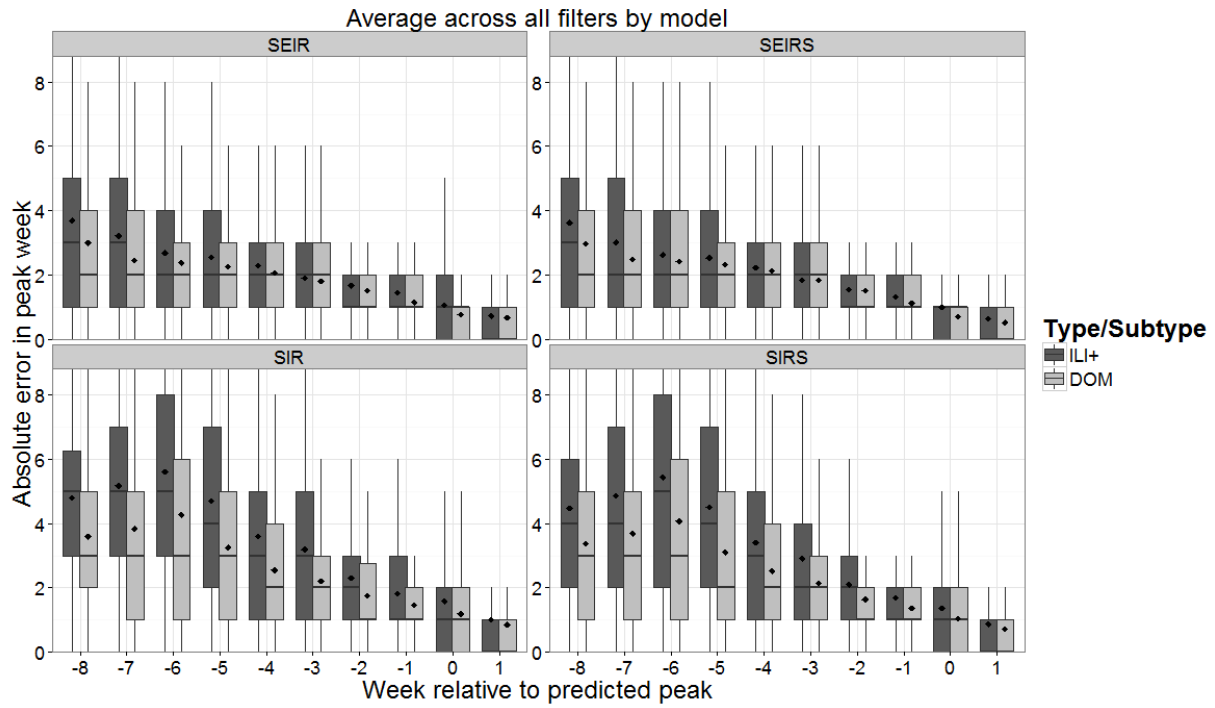
Web Figure 3. As for Figure 2, but stratified by filter type.

Web Figure 4. As for Figure 2, but stratified by model type.

Web Figure 5. RMSE error over a four-week horizon (RMSE 4) for the ILI+ (x-axis) and SUM (y-axis) forecasts. Each subplot depicts a different peak week lead time. Negative values indicate mean trajectory forecast of a future peak. Points below the diagonal indicate an improvement with use of the SUM (type/subtype-specific) forecasts. Relative error, defined as (ILI+ - SUM)/ILI+, is presented using the color-coding of individual points. The shape of the points indicated the number of types/subtypes used to generate SUM forecast (one vs. two or more types/subtypes). NS indicates no significant difference among the SUM and ILI+ approaches; a positive difference indicates significance at P <0.05.
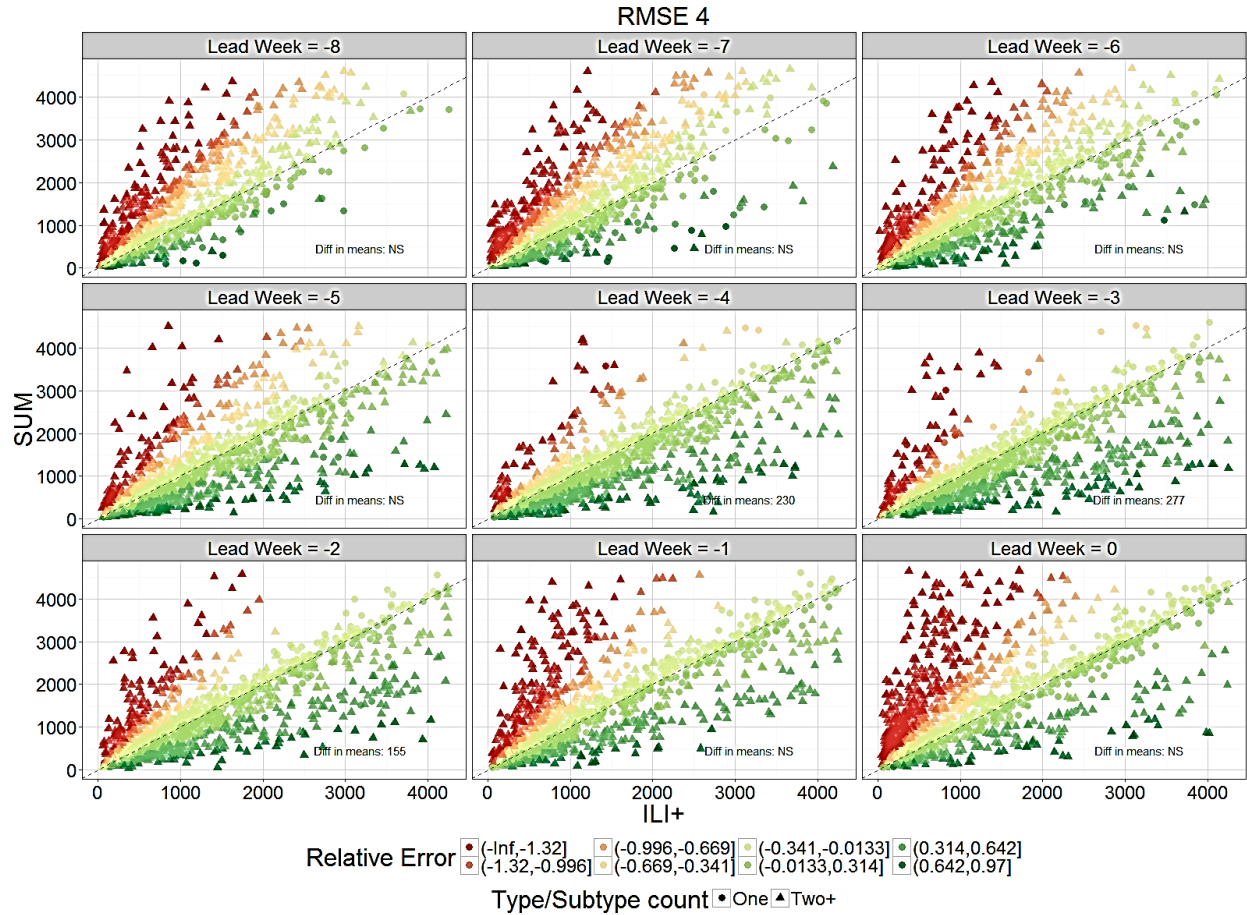
Web Table 1. Wilcoxon signed-rank test comparing predictions of ILI+ made following model training with ILI+ to predictions of ILI+ made as sum of H1N1+, H3N2+ and/or B+ predictions (SUM) broken down by season. The values in the table are a log of the Wilcoxon signed-rank test statistic, which is a measure of the distance of the median difference in errors from zero. NS indicates the difference is not statistically significant (p > 0.01). All numbers indicate lower error for the SUM forecasts. The types/subtypes predominantly in circulation each season are shown parenthetically.

|	2003-2004 (H3N2, B)	2004-2005 (H3N2, B)	2005-2006 (H1N1, B)	2006-2007 (H1N1, B)	2007-2008 (All 3)	2010-2011 (H3N2, B)	2011-2012 (H3N2, B)	2012-2013 (All 3)	2013-2014 (H3N2, B)	2014-2015 (H3N2, B)
Peak Week Correct (%)	NS	NS	NS	NS	NS	12.49	11.16	11.40	NS	NS
Peak Week Error (%)	NS	NS	NS	NS	11.82	12.54	11.71	12.65	NS	NS
Peak Intensity Correct (%)	NS	NS	NS	NS	NS	12.54	12.34	11.00	NS	10.22
Peak Intensity Error (%)	NS	NS	NS	NS	14.25	14.39	14.38	13.05	NS	NS

Web Table 2. Wilcoxon signed-rank test comparing predictions of ILI+ made following model training with ILI+ to predictions of ILI+ made as sum of H1N1+, H3N2+ and/or B+ predictions (SUM). The values in the table are a log of the Wilcoxon signed-rank test statistic, which is a measure of the distance of the median difference in errors from zero. NS indicates the difference is not statistically significant (p > 0.01). All numbers indicate lower error for the SUM forecasts. The first column evaluates differences over all forecasts; the second evaluates differences for SUM forecasts for the 95 cities only; the third column evaluates differences for SUM forecasts for the 50 states only.

	All SUM forecasts	City Only	State Only
Peak Week Correct	15.28	14.50	13.02
Peak Week Error	15.79	15.05	13.46
Peak Intensity Correct	16.61	15.94	14.09
Peak Intensity Error	18.89	18.09	16.66

Web Figure 6. The mean infectious period, D, estimated when the effective reproductive number is maximal using the ensemble adjustment Kalman filter and 4 different dynamic models. Estimates are shown for each season and for H1N1+ (yellow), H3N2+ (red), ILI+ (blue) and B+ (green). The diamond denotes the mean and the dot denotes the median, the whiskers demarcate the interquartile range.
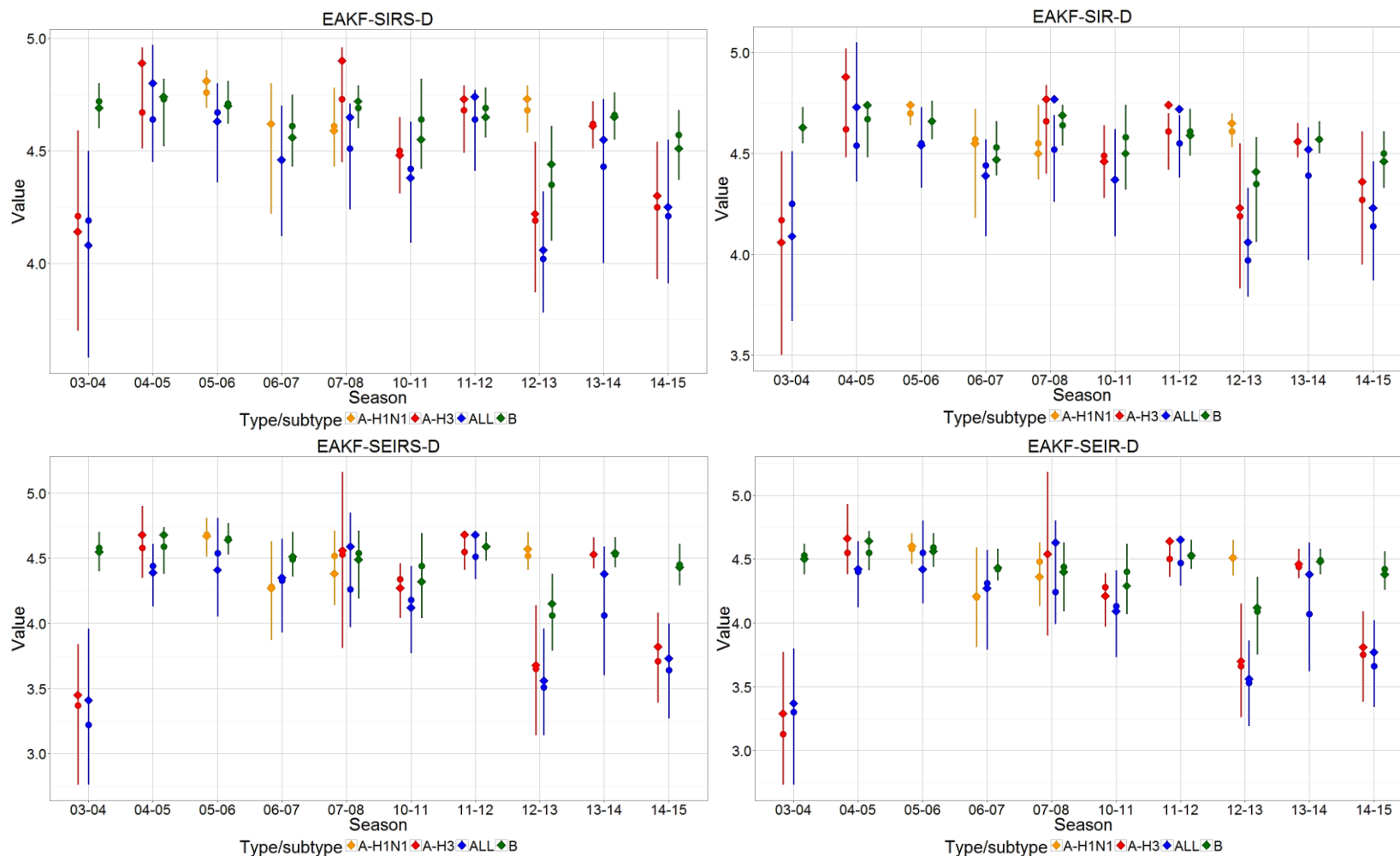
Web Figure 7. The basic reproductive number, R_0, estimated when the effective reproductive number is maximal using the ensemble adjustment Kalman filter and 4 different dynamic models. Estimates are shown for each season and for H1N1+ (yellow), H3N2+ (red), ILI+ (blue) and B+ (green). The diamond denotes the mean and the dot denotes the median, the whiskers demarcate the interquartile range.
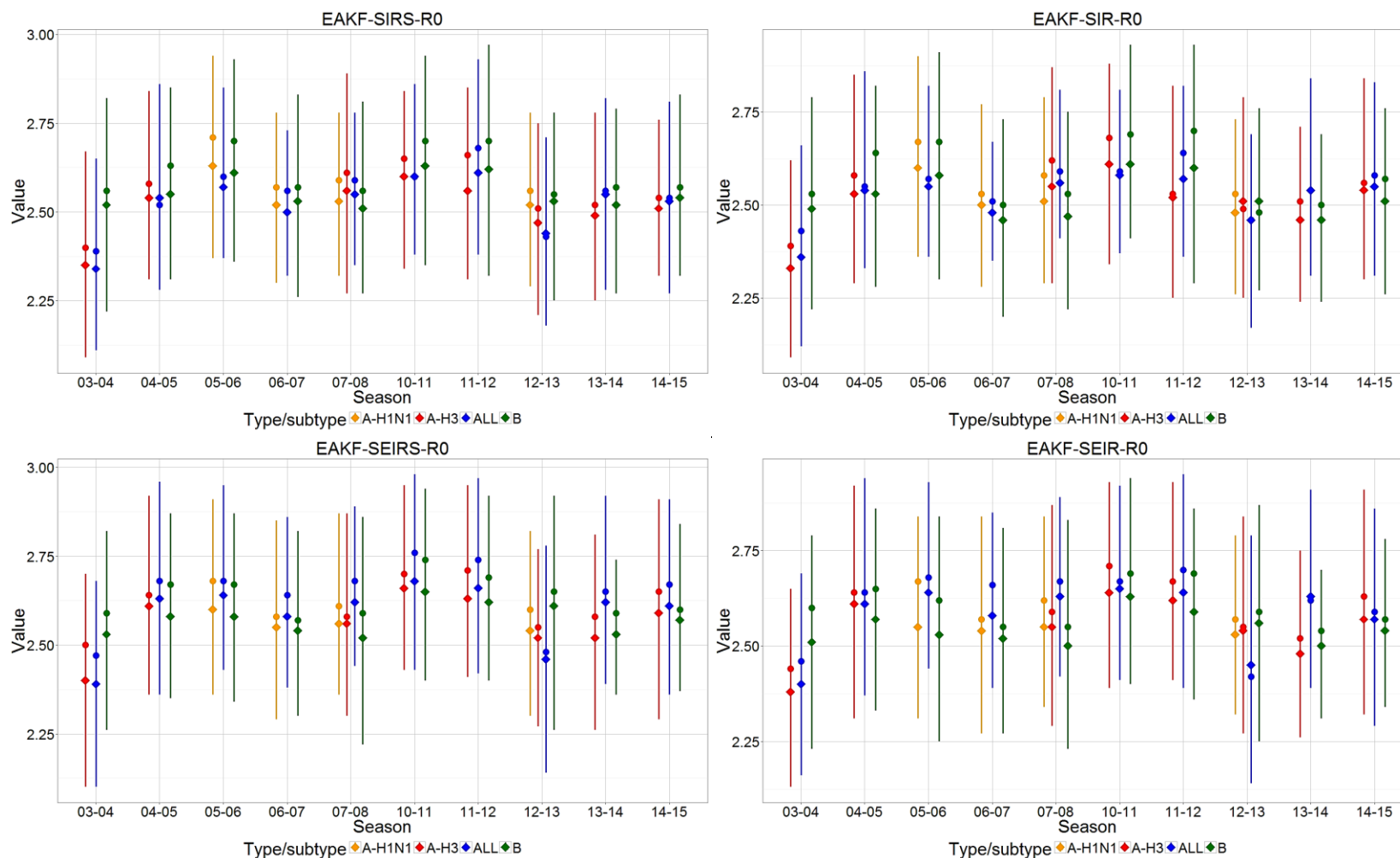
Web Figure 8. The effective reproductive number, R_e, estimated when the effective reproductive number is maximal using the ensemble adjustment Kalman filter and 4 different dynamic models. Estimates are shown for each season and for H1N1+ (yellow), H3N2+ (red), ILI+ (blue) and B+ (green). The diamond denotes the mean and the dot denotes the median, the whiskers demarcate the interquartile range.