Weighted averaging of recent survey indices

Raymond A. Webster

Introduction

In recent years, the International Pacific Halibut Commission (IPHC) has used an equal weighting of the three most recent years' setline survey index values to apportion the estimated biomass among regulatory areas. However, this method was somewhat arbitrary and we wished to develop a more statistically defensible approach. Here we explore weighting of annual survey index values from a statistical perspective, in which the relative variability within a year and between years determines how much weight should be given to earlier data.

Methods

Kalman Filter

The Kalman Filter is a method that efficiently updates an estimate of a measured quantity to incorporate the most recent observations. Its usefulness is greatest when there are large amounts of data (temporal, or spatio-temporal) as the updating does not require matrix computations using the full historical dataset. That is, if we wish to estimate θ_t, where t is a time index, an estimate is obtained from observations Y_1, ..., Y_{t-1}, and then updated by addition of a term that is a function of the most recent data, Y_t, only.

Suppose at time t we observed the quantity Y_t, a measure of some underlying variable θ_t. For our purposes, the measure is assumed to be unbiased, but made with error, i.e.,

$$Y_t = \theta_t + \varepsilon_t \tag{1}$$

where ε_t is the observation error, with $\varepsilon_t \sim N\left(0, \sigma_t^2\right)$. The second part of the model introduces a dynamic system equation describing how the unobserved process θ_t evolves with time (Meinhold and Singpurwalla, 1983):

$$\theta_t = G_t \theta_{t-1} + \eta_t \tag{2}$$

G_t describes our understanding of how θ_t changes with time, and in Kalman filtering is assumed to be known, and η_t is a random error term, $\eta_t \sim N\left(0, \gamma_t^2\right)$. Here we assume that $\gamma_t^2 = \gamma^2$, so does not change with time. The Kalman filter works iteratively, by combining our knowledge of θ_t from past information obtained using Equation 2 with the most recent data Y_t. This produces an estimate of θ_t that is the weighted average of current data, Y_t, and past observations, Y_1, ..., Y_{t-1}, in which the weights depend on the relative variability of the measurement and dynamic processes, Equations 1 and 2 respectively; if Y_t is a precise estimate of θ_t, then relatively little weight will be given to past observations. Explicit expressions for the weights are given in the Appendix.

While the usual Kalman filter approach to estimation of θ_t assumes that G_j (j=1,...,t), the variance parameters, and the initial value of the underlying process, θ_0, are known, alternative methods allow these parameters to be estimated within the modelling. In particular, in a fully Bayesian approach, we can put priors on these parameters and determine their posterior distributions, along with the posterior distributions of the θ_t. For large data sets, this will not have the advantage of computational efficiency that the recursive Kalman filter updating has, but in our application to time series of survey WPUE, the small size of the data sets means this is not a concern. Under a Bayesian approach, the estimate of each θ_t is no longer simply a weighted average of past data, since each depends on the value of variance parameter, τ^2, which depends on the entire series of observed data. This is a disadvantage if, in addition to improving the estimation of θ_t, obtaining explicit weights themselves is of interest. However, approximate weights could still be obtained by substituting the estimates of G_j and θ_0 into the expressions above.

Variability and weights

The Kalman filter weights depend on the relative sizes of the two sources of variability: within year and between year. If the between year variance of the process, γ^2, is large relative to measurement variance within a year (σ_j^2, j=1,...,t), then recent observations will have more weight than if γ^2 were small relative to σ_j^2. In other words, if a process deviates relatively little from the overall trend described by G_j, greater weight will be placed on past observations than if there is a lot of variability about the trend, in which case only recent data will affect the current estimate. We can illustrate this through some examples where, for simplicity, we suppose $\sigma_j^2 = \sigma^2$ (j=1,...,t) so that measurement variance does not change with time. Let $R = \gamma/\sigma$, the ratio of the two standard deviations. Results for $G_j = 1$, 0.95, or 1.05 (j=1,...,t) are shown in Tables 1-3 respectively. Note that the weights of the data, Y_j, do not sum to 1 unless $G_j = 1$, because (from Equation 2) θ_t has conditional expectation $G_t\theta_{t-1}$, and so the G_j affect the weights of the data: if all $G_j < 1$, the decreasing trend leads to weights that sum to a value less than 1, while $G_j > 1$ (increasing trend) leads to weights that sum to more than 1 (see Appendix). The tables show that as the ratio of process and measurement variance increases, the most recent data get greater weight. Therefore, if the survey produces a relatively precise estimate each year, but the true mean can change by a large amount between years, little weight will be given to past observations. Conversely, if the survey estimate is imprecise and the process changes little from year to year, past observations are quite informative about the current state of the process and receive greater weight.

Application to IPHC survey WPUE time series

Let Y_t be the survey WPUE for a regulatory area measured in year *t*. This is assumed to be an unbiased estimate of an underlying "true" WPUE (θ_t here, as per Equation 1) something we could obtain in theory by saturating an area with survey stations. The θ_t in turn are realisations of the dynamic process defined by Equation 2.

With a survey WPUE time series, a functional form for G_t may not be obvious. If WPUE is steadily decreasing or increasing, then a constant G_t not equal to 1 is reasonable, but over a long enough time series, a more complex process will be apparent. In their use of the Kalman filter to compute weights for averaging survey indices of Pacific cod, Thompson and Dorn (2004) assume that $G_t = 1$, and thus Equation 2 represents a trendless random walk. The random walk assumption seems unreasonable for the IPHC survey time series, in which clear upward or downward trends occur over periods of several years. However, such a model may still fit well depending on the period of observation, as random walks will track up or down by chance.

Another approach to dealing with a trend would be to remove it prior to using the Kalman filter by fitting some kind of regression model. This would lead to a Kalman filter estimate of θ_t that would then require the trend to be added back to produce an estimate of the survey index at time *t*. An alternative would be to directly estimate G_t within the model using the more fully Bayesian methods discussed above. In this work we consider linear and quadratic trends in G_t with time.

Bayesian modelling

We fit models using Markov chain Monte Carlo (MCMC) algorithms in WinBUGS (Spiegelhalter et al. 2003) in order to produce improved estimate of the most recent *WPUE* value for each area. Vague Gaussian priors, specifically $N(0,10000)$, are placed on the initial value, θ_0, and the parameters of the model for G_t. A non-informative Uniform(0,300) prior is placed on γ, the square root of the variance component γ^2. To simplify the modelling, the variance terms σ_t^2 are not estimated within the modelling, but calculated from the station data within each regulatory area for each year, thus avoiding having to work with the much larger station-level data set in the Bayesian modelling.

We considered four models for the G_t:

- $G_t = 1$, the trendless random walk
- $G_t = b_0$, constantly increasing or decreasing *WPUE*
- $G_t = b_0 + b_1 t$, linear model allowing changes in *WPUE* to accelerate or decelerate.
- $G_t = b_0 + b_1 t + b_2 t^2$, quadratic model

The Deviance Information Criterion (*DIC*, Spiegelhalter et al., 2002) was used to select among the models. The "best" fitting model is that with lowest DIC. Survey data are from 1998-2009, and our goal is to use the modelling to improve the estimate of 2009 *WPUE*.

Results

DIC values were generally similar for all but the quadratic model (Table 4), with models $G_t = 1$ and $G_t = b_0$ each being selected for four regulatory areas. (For Area 3A, these two models had the same *DIC*, and we selected the less complex of the two.) For the four areas for which the constant trend model, $G_t = b_0$, was selected, estimates of the trend parameter ranged from 0.85 for Area 2A to 0.92 for Areas 3B and 4B. Thus for these areas, we estimate that on average, each year's WPUE is 85-92% of the preceding year's WPUE. However, $G_t = 1$ does not imply no change in WPUE. As mentioned above, a trendless random walk can also have increases or decreases in WPUE consistent with what we observe in the IPHC survey data.

For most areas, estimates of *WPUE* in 2009 ($\hat{\theta}_t$) from the modelling were very close to the observed values (Table 5). This is because estimates of the process standard deviation, γ, are high relative to observation error, as shown by high values of the ratio R in Table 5. As seen in the earlier examples, this will lead to high weights for the current year's data (i.e., observed 2009 *WPUE*) and low weight for preceeding years, as confirmed by the estimated weights in Table 6. Only Areas 2A and 4A have *WPUE* estimates from the modelling that differ by more than a small amount from the observed values: Area 2A's estimate of 9.1 is 11% higher than the observed value of 8.0, and Area 4A's estimate of 78.0 is 6.5% less than the observed *WPUE*. These areas have the lowest ratios of process to observation variability (*R*, Table 5), and therefore have higher weights for past observations. In the case of Area 2A, observed *WPUE* in previous years has been higher than in 2009, and therefore past data have an upward influence on the model estimate of *WPUE*. For Area 4A, the low value in 2007 appears to have led to the lower estimate of 2009 *WPUE* from the modelling.

By using additional information in estimating the most recent *WPUE* value, we might also expect to improve the precision of our estimate. Improvements turn out to be very small; standard deviations for the model estimates are only slightly lower than those computed from the observed data only (Table 5). This is because relatively little weight is given to past observations, and therefore little additional information is used in improving the estimate of *WPUE*.

Selecting weights for all areas

The weights in Table 6 differ among areas, and depending on the selected model, may not sum to one. In choosing weights to use in practice, it is desirable from a management point of view to have the same weights for all regulatory areas and to keep them constant over time, and (for ease of interpretation by the public) have weights that sum to one. Only the random walk model has weights that sum to one. Other models' weights depend on the model for G_t, which will also have to be re-estimated each year as new data become available. From Table 4, when the $G_t = b_0$ was selected, the fit of the random walk model as measured by the *DIC* was never much worse. Given that over longer periods of time, $G_t = b_0$ (steady increase or decrease) will be an implausible model of changes in *WPUE*, and in the absence of a better model describing how *WPUE* changes with time, we are justified in using weights from the random walk models as a basis for defining coastwide weights.

In Table 7 we present estimated weights from the random walk model for each area for estimating the mean *WPUE* in 2009 and 2010 (the 2010 data were not available when we did the original analysis described above). Very little weight (0 to 0.008%) is given to data more than two years old for any area in either year, and so using non-zero weights for only the three most recent years' data is appropriate. The most weight given to past observations is for Areas 2B and 2A in 2010, for which (respectively) 20.2% and 24.1% of the weight is given to data prior to 2010. In choosing weights to apply to all areas, it may be preferable to give too much weight to data from the past two years than too little; in those years in which the survey data are most variable relative to variability among years, the additional information from previous years can improve estimation of the *WPUE* index, and we would prefer to err on the side of a little too much stability in the index than too little. Therefore, based on values in Table 7, a weighting of 75% for the current year, 20% for the previous year, and 5% for two years previous, would seem a reasonable choice as a set of coastwide weights.

Discussion

Previously the IPHC has weighted the past three years' observations equally when using area *WPUE* for apportionment. One justification for doing this is that the observation error in a single year's survey index means the true value may differ somewhat from the observed value, and including past observations in the index will protect against what could just be changes due to random sampling variation. As we have shown here, for most areas observation error is relatively small, and it is apparent that equal weighting will lead to a biased index, with too much influence given to past observations that are often quite different from the most recent data. An alternative 2:2:1 (recent data first) has also been proposed, but from the perspective of bias and variance, this too will weight past data too highly.

Rapid changes in catches in response to the changing index may be undesirable from the perspective of the fishers, who seek some degree of stability and predictability in the annual catch limits. The three-year equal weighting works to provide that by buffering changes. In doing so it is operating in a somewhat similar manner to the "slow up, fast down" (SUFD) policy. However, the weighting acts only on apportionment, so the slowing down of changes is only relative to other areas. This is not true of SUFD, which is applied independently to each individual area. Nevertheless, such buffering should not be the function of a weighting method, and is better considered separately as a management tool if that is desired.

The three-year equal weighting had no scientific basis, and it is very difficult to scientifically defend a weighting scheme which says that data from two years ago are as relevant to this year's biomass index as data that were actually collected this year. A weighting approach based on the analyses we have presented here instead has the intention of giving past data a weight that is appropriate for their relevance to today's survey index.

Appendix: Kalman filter weight formulae

Let $\tau_\theta = \gamma^{-2}$ and $\tau_{Y_j} = \sigma_j^{-2}$ and define the following

$$\beta_1 = \tau_\theta$$

$$\beta_j = \frac{\tau_{j-1}\tau_\theta}{\tau_{j-1} + G_j^2 \tau_\theta}, \quad j = 2, \ldots, t$$

$$\tau_j = \tau_{Y_j} + \beta_j, \quad j = 1, \ldots, t$$

$$w_{\theta_j} = \frac{G_j \beta_j}{\tau_{Y_j} + \beta_j}, \quad j = 1, \ldots, t$$

$$w_{Y_j} = \frac{\tau_{Y_j}}{\tau_{Y_j} + \beta_j}, \quad j = 1, \ldots, t$$

Then the Kalman filter estimate of θ_t is given by

$$\hat{\theta}_t = \sum_{j=1}^{t} \omega_t Y_t + \omega_0 \theta_0$$

where

$$\omega_j = w_{Y_j} \prod_{i=j+1}^{t} w_{\theta_i}, \quad j = 1, \ldots, t-1$$

$$\omega_t = w_{Y_t}$$

and $\omega_0 = \prod_{i=1}^{t} w_{\theta_i}$.

References

Meinhold, R. J. and Singpurwalla, N. D. (1983). Understanding the Kalman filter. American Statistician, 37, 123-127.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A. (2002). Bayesian measures of model complexity and fit. (B).Journal of the Royal Statistical Society, 64, 583–640.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003). WinBUGS Users Manual, Version 1.4.

Thompson, G. G. and Dorn, M. W. (2004). Assessment of the Pacific cod stock in the eastern Bering Sea and Aleutian Islands area. *In* Plan Team for Groundfish fisheries of the Bering Sea and Aleutian Islands, Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands regions, pp 185-382. North Pacific Fishery Management Council, Anchorage, AK.

Table 1. Weights of Y_j and θ_0 for examples with $G_j = 1$, j=1,...,t=10.

j	$R = \gamma / \sigma$			
	0.5	1	2	5
t=10	0.390	0.618	0.828	0.963
9	0.238	0.236	0.142	0.036
8	0.145	0.090	0.024	0.001
7	0.088	0.034	0.004	0.000
6	0.054	0.013	0.001	0.000
5	0.033	0.005	0.000	0.000
4	0.020	0.002	0.000	0.000
3	0.012	0.001	0.000	0.000
2	0.006	0.000	0.000	0.000
1	0.003	0.000	0.000	0.000
0	0.011	0.000	0.000	0.000
Sum	1.000	1.000	1.000	1.000

Table 2. Weights of Y_j and θ_0 for examples with $G_j = 0.95$, j=1,...,t=10.

j	$R = \gamma / \sigma$			
	0.5	1	2	5
t=10	0.368	0.608	0.826	0.963
9	0.221	0.227	0.137	0.034
8	0.133	0.084	0.023	0.001
7	0.080	0.031	0.004	0.000
6	0.048	0.017	0.001	0.000
5	0.029	0.004	0.000	0.000
4	0.017	0.002	0.000	0.000
3	0.010	0.001	0.000	0.000
2	0.005	0.000	0.000	0.000
1	0.002	0.000	0.000	0.000
0	0.009	0.000	0.000	0.000
Sum	0.921	0.969	0.990	0.998

Table 3. Weights of Y_j and θ_0 for examples with $G_j = 1.05$, j=1,...,t=10.

j	$R = \gamma/\sigma$			
	0.5	1	2	5
t=10	0.414	0.629	0.831	0.963
9	0.255	0.245	0.147	0.037
8	0.157	0.096	0.026	0.001
7	0.096	0.037	0.005	0.000
6	0.059	0.015	0.001	0.000
5	0.036	0.006	0.000	0.000
4	0.022	0.002	0.000	0.000
3	0.013	0.001	0.000	0.000
2	0.007	0.000	0.000	0.000
1	0.003	0.000	0.000	0.000
0	0.014	0.000	0.000	0.000
Sum	1.076	1.030	1.010	1.002

Table 4. *DIC* values comparing the fit of alternative trend models.

Area	Model			
	$G_t = 1$	$G_t = b_0$	$G_t = b_0 + b_1 t$	$G_t = b_0 + b_1 t + b_2 t^2$
4D	**94.2**	94.7	94.6	95.5
4B	101.5	**101.3**	101.9	105.2
4A	112.3	**111.2**	112.0	114.3
3B	100.8	**100.2**	100.6	102.1
3A	**102.7**	102.7	103.4	102.9
2C	**103.3**	103.5	104.0	106.8
2B	**87.9**	88.1	88.4	90.6
2A	69.2	**67.8**	69.2	70.5

Table 5. Posterior means and standard deviations of setline survey WPUE for selected model for each area in 2009.

Area	Y_t (WPUE)	$\hat{\sigma}_t$ (SD)	model	$\hat{\theta}_t$	SD	$\hat{b}_0$ (SD)	$\hat{\gamma}$ (SD)	$R=\hat{\gamma}/\hat{\sigma}_t$
4D	69.2	11.8	$G_t=1$	69.0	11.3	0.89 (0.09)[1]	42.6 (16.8)	3.6
4B	106.8	11.9	$G_t=b_0$	103.3	11.1	0.92 (0.06)	24.8 (8.9)	2.1
4A	83.4	10.9	$G_t=b_0$	78.0	10.5	0.89 (0.04)	20.3 (14.3)	1.9
3B	113.0	4.9	$G_t=b_0$	113.2	4.8	0.92 (0.04)	34.7 (20.3)	7.1
3A	148.8	6.5	$G_t=1$	150.2	6.5	0.96 (0.05)[1]	38.5 (13.1)	5.9
2C	115.6	7.2	$G_t=1$	115.2	6.9	0.95 (0.04)[1]	29.2 (9.6)	4.1
2B	86.3	5.6	$G_t=1$	85.5	5.3	0.98 (0.06)[1]	13.7 (4.6)	2.4
2A	8.0	2.3	$G_t=b_0$	9.1	2.1	0.85 (0.07)	3.8 (3.4)	1.7

1. Values estimated using model $G_t=b_0$ and included for comparison with other regulatory areas.

Table 6. Approximate weights for data when estimating 2009 WPUE using the Kalman filter with the selected model (Table 5). Estimates from the Bayesian modelling are used for G_t, γ, and θ_0.

Year	4D	4B	4A	3B	3A	2C	2B	2A
2009	0.932	0.837	0.803	0.981	0.973	0.946	0.879	0.838
2008	0.065	0.119	0.140	0.017	0.026	0.052	0.095	0.062
2007	0.003	0.024	0.027	0.000	0.001	0.002	0.023	0.020
2006	0.000	0.003	0.003	0.000	0.000	0.000	0.002	0.020
2005	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.004
2004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003
2003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003
2002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1999	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
1998	-	0.000	0.000	0.000	0.000	0.000	0.000	-
1997	-	0.000	0.000	0.000	0.000	0.000	0.000	-

Table 7. Approximate weights for data when estimating 2009 and 2010 WPUE using the Kalman filter with the random walk model. Estimates from the Bayesian modelling are used for G_t, γ, and θ_0.

Year	4D	4B	4A	3B	3A	2C	2B	2A
2010	0.977	0.950	0.940	0.994	0.981	0.940	0.798	0.759
2009	0.022	0.044	0.056	0.006	0.018	0.057	0.176	0.224
2008	0.002	0.005	0.004	0.000	0.001	0.004	0.021	0.011
<2008	0.000	0.001	0.000	0.000	0.000	0.000	0.007	0.005
2009	0.932	0.872	0.941	0.992	0.973	0.946	0.879	0.937
2008	0.065	0.107	0.055	0.008	0.026	0.052	0.095	0.045
2007	0.003	0.019	0.003	0.000	0.001	0.002	0.023	0.011
<2007	0.000	0.002	0.000	0.000	0.000	0.000	0.002	0.008