J Geod (2013) 87:461–473
DOI 10.1007/s00190-013-0623-6

ORIGINAL ARTICLE

BLUE, BLUP and the Kalman filter: some new results

P. J. G. Teunissen · A. Khodabandeh

Received: 26 October 2012 / Accepted: 2 February 2013 / Published online: 23 February 2013

Abstract In this contribution, we extend 'Kalman-filter' theory by introducing a new BLUE–BLUP recursion of the partitioned measurement and dynamic models. Instead of working with known state-vector means, we relax the model and assume these means to be unknown. The recursive BLUP is derived from first principles, in which a prominent role is played by the model's misclosures. As a consequence of the mean state-vector relaxing assumption, the recursion does away with the usual need of having to specify the initial state-vector variance matrix. Next to the recursive BLUP, we introduce, for the same model, the recursive BLUE. This extension is another consequence of assuming the state-vector means unknown. In the standard Kalman filter set-up with known state-vector means, such difference between estimation and prediction does not occur. It is shown how the two intertwined recursions can be combined into one general BLUE–BLUP recursion, the outputs of which produce for every epoch, in parallel, the BLUP for the random state-vector and the BLUE for the mean of the state-vector.

Keywords Best linear unbiased estimation (BLUE) · Best linear unbiased prediction (BLUP) · Minimum mean squared error (MMSE) · Misclosures · Kalman filter · BLUE–BLUP recursion

P. J. G. Teunissen · A. Khodabandeh (✉)
Department of Spatial Sciences, GNSS Research Centre,
Curtin University of Technology, Perth, Australia
e-mail: amir.khodabandeh@curtin.edu.au

P. J. G. Teunissen
Department of Geoscience and Remote Sensing,
Delft University of Technology, Delft, The Netherlands
e-mail: p.teunissen@curtin.edu.au

1 Introduction

To determine best estimators or best predictors, the minimum mean squared error (MMSE) criterion is often used. Different MMSE predictors exist however. They depend on the class of functions for which the MMSE principle is applied. Examples of different MMSE predictors are the conditional mean as best predictor (BP), the best linear predictor (BLP), the best integer equivariant predictor (BIEP), or the best linear unbiased predictor (BLUP), see e.g. Goldberger (1962), Anderson and Moore (1979), Stark and Woods (1986), Sanso (1986), Simon (2006), Teunissen (2007). Although the same principle is applied, these MMSE predictors all have different performances.

In the literature, the Kalman filter is derived as either a BP or a BLP, see e.g. Kalman (1960), Gelb (1974), Kailath (1981), Candy (1986), Brammer and Siffling (1989), Jazwinski (1991), Gibbs (2011). Both these predictors, BP and BLP, require the mean of the to-be-predicted random vector to be known. This is why in the derivation of the Kalman filter one usually assumes the mean of the random initial state-vector to be known, see for instance the contributions by Sorenson (1966, p. 222), Maybeck (1979, p. 204), Anderson and Moore (1979, p. 15), Stark and Woods (1986, p. 393), Bar-Shalom and Li (1993, p. 209), Kailath et al. (2000, p. 311), Simon (2006, p. 125), and Grewal and Andrews (2008, p. 138).

Despite the BLP approach, it is indeed sometimes acknowledged that the mean of the initial state-vector is not known. The approach then taken is to treat the initial state-vector as being *diffuse*, meaning that its variance matrix tends to infinity, see e.g. Harvey and Phillips (1979), Ansley and Kohn (1985), de Jong (1991). The proposed approach in practice is then to initialize the Kalman filter with a sufficiently 'large' variance matrix. With such an approach, however, the Kalman filter is still derived and presented within the BLP context.

We believe that the BLP derivation of the Kalman filter is not appropriate in case the mean of the random state-vector is unknown, a situation that applies to many, perhaps even most, engineering applications. Not the BLP principle, but the BLUP principle should be applied in case the mean state-vectors are unknown. In this contribution, we derive and present the recursive BLUP from first principles and show how such an approach does away with the need to assume the mean and variance matrix of the initial state-vector to be known.

Next to the recursive BLUP, we also present, for the same model, the recursive best linear unbiased estimator (BLUE). To appreciate the difference with the recursive BLUP, it is important to make a sharp distinction between prediction and estimation. We speak of prediction if observables are used to guess the outcome of a random vector and we speak of estimation if the observables are used to guess the value of an unknown nonrandom vector.

Our development of the BLUE–BLUP recursion is an extention of standard 'Kalman filter' theory. This extension is a consequence of our relaxing assumptions that the means of the random state-vectors are unknown. Since these means are assumed unknown, the problem of estimation can be addressed as well. In the standard Kalman filter set-up with known state-vector means, this difference between estimation and prediction does not occur since one is then only left with BLP instead of with BLUP of the state-vectors.

This contribution is organized as follows. In Sect. 2, we briefly review the necessary ingredients of prediction and estimation for linear models. Essential in our presentation is the role given to the misclosures of the linear model. We treat prediction and estimation on an equal footing and show how predictors and estimators are driven by the way misclosures are mapped. We also show how the BLUE and BLUP can be decomposed into misclosures and any LUE or LUP. This decomposition forms the basis for our development of the BLUE–BLUP recursion in later sections.

In Sect. 3, we present the general BLUE–BLUP measurement update equations for a time series of vectorial observables. Through a one-to-one mapping, it is shown how the sequential prediction errors of the misclosures form the basis of the predicted residuals. No assumptions are here yet made on the structure of the variance–covariance matrices.

In Sect. 4, we develop the BLUP recursion for the partitioned measurement and dynamic models that form the basis of the standard Kalman filter. Instead of the standard assumption of known mean state-vectors, we assume the means of the random state-vectors to be unknown. Through this relaxation the initialization issue gets resolved, whereby it is shown that the variance matrix of the initial state-vector is not needed anymore. In Sect. 5, we extend standard 'Kalman-filter' theory further by introducing, next to the BLUP recursion, the BLUE recursion of the means of the random state-vectors. It is shown how the two recursions are intertwined and how their difference is driven by the presence of system noise. Finally, we show how the two recursions can be combined into one general BLUE–BLUP recursion. It outputs for every epoch, in parallel, the BLUP for the random state-vector and the BLUE for the mean of this state-vector.

We make use of the following notation: we use the underscore to denote a random vector. Thus $\underline{x}$ is random, while x is not. $\mathsf{E}(\cdot)$ and $\mathsf{D}(\cdot)$ denote the expectation and dispersion operator, while $\mathsf{C}(\cdot,\cdot)$ denotes the covariance operator. The norm of a vector is denoted as $||\cdot||$. Thus $||\cdot||^2 = (\cdot)^T(\cdot)$.

2 Estimation and prediction in linear models

2.1 Estimation and prediction

As our point of departure, we take the following linear model of observation equations

$$\underline{y} = Ax + \underline{e} \tag{1}$$

with mean and dispersion

$$\mathsf{E}(\underline{e}) = 0\ ,\ \mathsf{D}(\underline{y}) = \mathsf{D}(\underline{e}) = Q_{yy}$$

where $\underline{y} \in \mathbb{R}^m$ is the random vector of observables and $x \in \mathbb{R}^n$ is the nonrandom vector of unknown parameters. The known matrix A, of order $m \times n$, is assumed to be of full column rank, and the variance matrix Q_{yy} is assumed to be positive definite.

The aims are to *estimate* a linear function of x, or, to *predict* the outcome of a random vector having a mean that is a linear function of x. The to-be-estimated linear function is given as

$$\bar{z} = A_z x, \tag{2}$$

with known $k \times n$ matrix A_z. The to-be-predicted random vector is given as

$$\underline{z} = A_z x + \underline{e}_z, \quad \text{with} \quad \mathsf{E}(\underline{e}_z) = 0 \tag{3}$$

This random vector is assumed to be stochastically related to $\underline{y}$. Their joint variance matrix is assumed given as

$$\mathsf{D}\left(\begin{bmatrix} \underline{y} \\ \underline{z} \end{bmatrix}\right) = \begin{bmatrix} Q_{yy} & Q_{yz} \\ Q_{zy} & Q_{zz} \end{bmatrix} \tag{4}$$

For our further development, the following canonical form of the linear model (1) is very helpful.

Lemma 1 (Linear model in canonical form) *Define the transformed vector of observables as* $[\underline{\hat{x}}^T, \underline{u}^T]^T = T\underline{y}$, *with transformation matrix* $T = [A^{+T}, B]^T$, *least-squares* *inverse* $A^+ = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$ *and where* B *is a basis matrix of the null space of* A^T, *i.e.* $A^T B = 0$. *Then*

$$\mathsf{E}\left(\begin{bmatrix}\underline{\hat{x}}\\ \underline{u}\end{bmatrix}\right) = \begin{bmatrix}I_n\\ 0\end{bmatrix} x, \quad \mathsf{D}\left(\begin{bmatrix}\underline{\hat{x}}\\ \underline{u}\end{bmatrix}\right) = \begin{bmatrix}Q_{\hat{x}\hat{x}} & 0\\ 0 & Q_{uu}\end{bmatrix} \tag{5}$$

Proof The mean and dispersion of (5) follow from an application of the mean and variance propagation laws to $T\underline{y}$. That the linear model (5) stands in a one-to-one relation with the original linear model formulation (1) follows from the invertibility of $T = [A^{+T}, B]^T \Leftrightarrow T^{-1} = [A, B^{+T}]$, with least-squares inverse $B^+ = (B^T Q_{yy} B)^{-1} B^T Q_{yy}$. □

With the above canonical form, the design matrix has the simple form $[I_n, 0]^T$, while the dispersion has become block diagonal. As $\underline{\hat{x}}$ and $\underline{u}$ stand in one-to-one correspondence with $\underline{y}$, while $\underline{\hat{x}}$ has the unity-matrix as design matrix and itself is uncorrelated with the zero-mean vector $\underline{u}$, one may expect that $\underline{\hat{x}}$ contains the full information for the determination of x. And indeed, $\underline{\hat{x}} = A^+\underline{y} \in \mathbb{R}^n$ is recognized as the BLUE of x, while $\underline{u} = B^T\underline{y} \in \mathbb{R}^{m-n}$ is recognized as the vector of *misclosures* (Teunissen 2000). The dimension of the vector of misclosures is equal to the redundancy, $r = m - n$, i.e. the vector of misclosures only exists in the presence of redundancy ($r > 0$). Note that the vector of misclosures is not unique, since for any invertible matrix L, the vector $\underline{v} = L\underline{u}$ is again a vector of misclosures, i.e. both B and BL^T are basis matrices of the null space of A^T.

The advantage of the decomposition $\underline{y} = T^{-1}[\underline{\hat{x}}^T, \underline{u}^T]^T$ is that it enables finding simple representations of estimators and predictors. The following lemma gives such representation for linear unbiased estimators/predictors of $\bar{z}$ and $\underline{z}$, respectively.

Lemma 2 (Linear unbiased estimation/prediction) *A linear function* $F\underline{y} + f_0$, *satisfying the condition* $\mathsf{E}(F\underline{y} + f_0) = A_z x$ *for all* x, *is called a linear unbiased estimator (LUE) of* $\bar{z} = A_z x$, *or a linear unbiased predictor (LUP) of* $\underline{z} = \bar{z} + \underline{e}_z$. *Any such LUE or LUP can be represented as*

$$\text{LUE/LUP} = A_z\underline{\hat{x}} + J\underline{u}, \; \textit{for some } J \in \mathbb{R}^{k\times r} \tag{6}$$

Proof With $\mathsf{E}(\underline{y}) = Ax$, the unbiasedness condition $\mathsf{E}(F\underline{y} + f_0) = A_z x, \forall x \in \mathbb{R}^n$, is fulfilled by setting $f_0 = 0$ and $FA = A_z$. Given the matrix equation $FA = A_z$, the general solution for F is given by the sum of a particular solution and the homogeneous solution. A particular solution is given by $A_z A^+$, while the homogeneous solution is provided by JB^T, for some matrix $J \in \mathbb{R}^{k\times r}$. Therefore, $F = A_z A^+ + JB^T$. Equation (6) follows then by substituting the result into $\text{LUE} = F\underline{y}$. □

The above representation shows that LUEs and LUPs differ only through their linear functions of the misclosure vector $\underline{u}$. Hence, it is through the choice of matrix J that specific LUEs and LUPs can be identified.

2.2 MMSE estimation/prediction and misclosures

To determine best estimators/predictors, the minimum mean squared error (MMSE) criterion is used. In case no restrictions are placed on the class of predictors, the best predictor of $\underline{z}$ in the MMSE sense is given by the conditional mean,

$$\text{BP} = \mathsf{E}(\underline{z}|\underline{y}) \tag{7}$$

[see e.g. Anderson and Moore (1979); Maybeck (1979); Teunissen et al. (2005)]. The BP is unbiased, but generally nonlinear, with exemptions, for instance in the Gaussian case. In case $\underline{y}$ and $\underline{z}$ are jointly Gaussian, the BP becomes linear and identical to the best linear predictor,

$$\text{BLP} = \bar{z} + Q_{zy}Q_{yy}^{-1}(\underline{y} - \bar{y}) \tag{8}$$

where $\bar{y}$ and $\bar{z}$ denote the mean of $\underline{y}$ and $\underline{z}$, respectively [see e.g. Bar-Shalom and Li (1993); Kailath et al. (2000); Teunissen (2008)]. Although the BLP is linear, it still requires knowledge of the means $\bar{y}$ and $\bar{z}$. These means are however unknown, since x is assumed unknown in case of the linear model (1). Instead of working within the unconstrained class of linear functions, we therefore work in the more restricted class as specified by the representation (6). To determine the best estimator/predictor within this class, use is made of the following lemma.

Lemma 3 (Minimum mean squared norm) *Let* $\underline{\varepsilon} \in \mathbb{R}^k$ *and* $\underline{u} \in \mathbb{R}^r$ *be given random vectors, with* $\mathsf{E}(\underline{u}) = 0$. *Then,*

$$\underline{\hat{\varepsilon}} = \underline{\varepsilon} - Q_{\varepsilon u}Q_{uu}^{-1}\underline{u} \tag{9}$$

has smallest mean squared norm within the class of random vectors $\underline{\varepsilon}_J = \underline{\varepsilon} + J\underline{u}$, $J \in \mathbb{R}^{k\times r}$, *i.e.*

$$\mathsf{E}||\underline{\hat{\varepsilon}}||^2 \le \mathsf{E}||\underline{\varepsilon}_J||^2 \;,\; \forall J \in \mathbb{R}^{k\times r} \tag{10}$$

Proof Since $\underline{\hat{\varepsilon}} = \underline{\varepsilon} - Q_{\varepsilon u}Q_{uu}^{-1}\underline{u}$ is uncorrelated with $\underline{u}$, we have the 'sum-of-squares' decomposition

$$\begin{aligned}\mathsf{E}||\underline{\varepsilon} + J\underline{u}||^2 &= \mathsf{E}||\underline{\varepsilon} - Q_{\varepsilon u}Q_{uu}^{-1}\underline{u} + (J + Q_{\varepsilon u}Q_{uu}^{-1})\underline{u}||^2\\ &= \mathsf{E}||\underline{\hat{\varepsilon}}||^2 + \mathsf{E}||(J + Q_{\varepsilon u}Q_{uu}^{-1})\underline{u}||^2\end{aligned}$$

from which (10) follows. □

This lemma shows that the mean squared norm of a random vector cannot be made smaller by adding uncorrelated linear functions of zero mean random vectors. We now use this lemma to determine the best estimator and the best predictor within the class of LUEs and LUPs, respectively.

Theorem 1 (BLUE/BLUP) *For any LUE of* $\bar{z}$, *with estimation error* $\underline{\varepsilon}_{\text{LUE}} = \bar{z} - \text{LUE}$, *and any LUP of* $\underline{z}$, *with prediction error* $\underline{\varepsilon}_{\text{LUP}} = \underline{z} - \text{LUP}$, *the BLUE and BLUP can be computed as*

$$\begin{bmatrix}\text{BLUE}\\ \text{BLUP}\end{bmatrix} = \begin{bmatrix}\text{LUE}\\ \text{LUP}\end{bmatrix} + Q_{\varepsilon u}Q_{uu}^{-1}\underline{u} \;\;,\;\; \underline{\varepsilon} = \begin{bmatrix}\underline{\varepsilon}_{LUE}\\ \underline{\varepsilon}_{LUP}\end{bmatrix} \tag{11}$$

Proof Consider a certain fixed LUP. Since the BLUP is a LUP, it follows from Lemma 2 that the BLUP can be written as BLUP = LUP + Ju for some J. Hence, their prediction errors, $\varepsilon_{\text{BLUP}} = \underline{z} -$ BLUP and $\varepsilon_{\text{LUP}} = \underline{z} -$ LUP, are related as $\underline{\varepsilon}_{\text{BLUP}} = \underline{\varepsilon}_{\text{LUP}} - J\underline{u}$. According to Lemma 3 (cf. 9), J must be chosen as $J = Q_{\varepsilon_{\text{LUP}},u} Q_{uu}^{-1}$ for $\mathsf{E}||\underline{\varepsilon}_{\text{BLUP}}||^2$ to be minimal. This proves the BLUP-part of (11). The proof of the BLUE part goes likewise. □

This result clearly shows the important role that is played by the vector of misclosures $\underline{u}$, both in best estimating $\bar{z}$ and in best predicting $\underline{z}$. In Sect. 5, we present the recursive counterpart of the above BLUE–BLUP expression.

The relations between the error covariance matrices are readily obtained from (11) through an application of the (co)variance propagation law. They are summarized in the following corollary.

Corollary 1 (Error covariance matrices) *The error covariance matrices of estimation and prediction satisfy:*

$$\begin{aligned} \mathsf{C}(\underline{\varepsilon}_{\text{BLUE}}, \underline{\varepsilon}_{\text{BLUP}}) &= \mathsf{C}(\underline{\varepsilon}_{\text{LUE}}, \underline{\varepsilon}_{\text{LUP}}) - Q_{\varepsilon_{\text{LUE}}u} Q_{uu}^{-1} Q_{u\varepsilon_{\text{LUP}}} \\ \mathsf{C}(\underline{\varepsilon}_{\text{BLUE}}, \underline{\varepsilon}_{\text{BLUP}}) &= \mathsf{C}(\underline{\varepsilon}_{\text{BLUE}}, \underline{\varepsilon}_{\text{LUP}}) = \mathsf{C}(\underline{\varepsilon}_{\text{LUE}}, \underline{\varepsilon}_{\text{BLUP}}) \\ \mathsf{C}(\underline{\varepsilon}_{\text{BLUE}}, \underline{u}) &= \mathsf{C}(\underline{\varepsilon}_{\text{BLUP}}, \underline{u}) = 0 \end{aligned} \tag{12}$$

This result shows that the misclosures are uncorrelated with both the best estimation errors and the best prediction errors.

2.3 From LUE to BLUE

As a direct consequence of the above theorem we have the following simple relationship between the BLUE and any LUE.

Corollary 2 (BLUE–LUE formula) *The BLUE and its variance matrix are related to that of any arbitrary LUE as*

$$\begin{aligned} \text{BLUE} &= \text{LUE} - Q_{\text{LUE},u} Q_{uu}^{-1} \underline{u} \\ Q_{\text{BLUE, BLUE}} &= Q_{\text{LUE, LUE}} - Q_{\text{LUE},u} Q_{uu}^{-1} Q_{u,\text{LUE}} \end{aligned} \tag{13}$$

Proof Since $\bar{z}$ is nonrandom, it follows from $\underline{\varepsilon}_{\text{LUE}} = \bar{z} -$ LUE that $\mathsf{C}(\underline{\varepsilon}_{\text{LUE}}, \underline{u}) = -\mathsf{C}(\text{LUE}, \underline{u})$. Substitution into (11) proves the first equation of (13). The second equation follows from an application of the variance propagation law, thereby making use of the fact that the misclosures are not correlated with the estimation error $\underline{\varepsilon}_{\text{BLUE}}$ (cf. 12) and thus also not with the BLUE itself, $\mathsf{C}(\text{BLUE}, \underline{u}) = 0$. □

This simple BLUE–LUE relation will be very useful in our later derivations of the recursive BLUE and the recursive BLUP. Here, three examples are given to see the BLUE–LUE relation at work.

Example 1 (Least-squares as LUE) The least-squares estimator LSE $= (A^TWA)^{-1}A^TW\underline{y}$ is a LUE of x. Hence, according to (13) the BLUE of x can be expressed in the LSE as

$$\text{BLUE} = \text{LSE} - Q_{\text{LSE},u} Q_{uu}^{-1} \underline{u} \tag{14}$$

which is identical to the LSE with $W = Q_{yy}^{-1}$.

Example 2 (Conditional adjustment) Consider the model of condition equations

$$B^T \mathsf{E}(\underline{y}) = 0 \tag{15}$$

It is the implicit formulation of the parametric model of observation equations $\mathsf{E}(\underline{y}) = Ax$. Since $\underline{y}$ is a LUE of $\mathsf{E}(\underline{y})$, the BLUE of $\mathsf{E}(\underline{y})$ can be written according to the BLUE–LUE formula (13) as $\hat{\underline{y}} = \underline{y} - Q_{yu} Q_{uu}^{-1} \underline{u}$. Substitution of $\underline{u} = B^T \underline{y}$ gives

$$\hat{\underline{y}} = \underline{y} - Q_{yy} B (B^T Q_{yy} B)^{-1} B^T \underline{y} \tag{16}$$

which is the BLUE of $\mathsf{E}(\underline{y})$ expressed in the design matrix B of the conditional model.

Example 3 (Tienstra's phased adjustment) Consider the partitioned model of condition equations

$$\begin{bmatrix} B_{[i-1]}^T \\ B_i^T \end{bmatrix} \mathsf{E}(\underline{y}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{17}$$

with the corresponding partitioned misclosure vector $\underline{u}_{[i]} = [\underline{u}_{[i-1]}^T, \ \underline{u}_i^T]^T$. Then, the BLUE of $\mathsf{E}(\underline{y})$ based on the first set of conditions $B_{[i-1]}^T \mathsf{E}(\underline{y})$, denoted as $\hat{\underline{y}}_{[i-1]}$, is a LUE of $\mathsf{E}(\underline{y})$ for the complete set of conditions. Hence, we have

$$\hat{\underline{y}}_{[i]} = \hat{\underline{y}}_{[i-1]} - Q_{\hat{y}_{[i-1]}u_{[i]}} Q_{u_{[i]}u_{[i]}}^{-1} \underline{u}_{[i]} \tag{18}$$

Since $\hat{\underline{y}}_{[i-1]}$ is uncorrelated with $\underline{u}_{[i-1]}$, it follows upon choosing the respresentation of the vector of misclosures as $\underline{u}_{[i]} = [\underline{u}_{[i-1]}^T, \ (B_i^T \hat{\underline{y}}_{[i-1]})^T]^T$ that

$$\hat{\underline{y}}_{[i]} = \hat{\underline{y}}_{[i-1]} - Q_{\hat{y}_{[i-1]}\hat{y}_{[i-1]}} B_i (B_i^T Q_{\hat{y}_{[i-1]}\hat{y}_{[i-1]}} B_i)^{-1} B_i^T \hat{\underline{y}}_{[i-1]} \tag{19}$$

This is Tienstra's formula for adjustment in phases (Tienstra 1956).

2.4 From LUP to BLUP

Similar to the BLUE–LUE relation, we have the following counterpart for prediction.

Corollary 3 (BLUP–LUP formula) *The BLUP, its error variance matrix and its variance matrix are related to that of any arbitrary LUP as*

$$\begin{aligned} \text{BLUP} &= \text{LUP} + Q_{z-\text{LUP},u} Q_{uu}^{-1} \underline{u} \\ Q_{\varepsilon_{\text{BLUP}}\varepsilon_{\text{BLUP}}} &= Q_{\varepsilon_{\text{LUP}}\varepsilon_{\text{LUP}}} - Q_{\varepsilon_{\text{LUP}},u} Q_{uu}^{-1} Q_{u,\varepsilon_{\text{LUP}}} \\ Q_{\text{BLUP, BLUP}} &= Q_{\text{LUP, LUP}} - Q_{\text{LUP},u} Q_{uu}^{-1} Q_{u,\text{LUP}} + Q_{zu} Q_{uu}^{-1} Q_{uz} \end{aligned} \tag{20}$$

Proof The first equation follows from (11), while the second and third follow from an application of the variance propagation law, thereby making use of the zero-covariance property (12). □

Note that $Q_{\varepsilon_{\text{BLUP}}\varepsilon_{\text{BLUP}}}$ and $Q_{\text{BLUP, BLUP}}$ are two different variance matrices. The first is the variance matrix of the prediction error $\underline{\varepsilon}_{\text{BLUP}} = \underline{z} - \text{BLUP}$, thus describing the MMSE quality of the predictor. The second does *not* describe the quality of prediction, but instead the precision of the predictor, i.e. the mean squared difference between the predictor and $\bar{z}$, instead of between the predictor and $\underline{z}$. In case of estimation, this difference is absent as we have $Q_{\varepsilon_{\text{BLUE}}\varepsilon_{\text{BLUE}}} = Q_{\text{BLUE, BLUE}}$ (cf 13).

From the above corollary we can also directly obtain the BLUP-BLUE relation.

Corollary 4 (BLUP-BLUE formula) *The BLUP, its error variance matrix and its variance matrix are related to that of the BLUE as*

$$\begin{aligned} \text{BLUP} &= \text{BLUE} + Q_{zu}Q_{uu}^{-1}\underline{u} \\ Q_{\varepsilon_{\text{BLUP}}\varepsilon_{\text{BLUP}}} &= Q_{z-\text{BLUE},z-\text{BLUE}} - Q_{zu}Q_{uu}^{-1}Q_{uz} \\ Q_{\text{BLUP, BLUP}} &= Q_{\text{BLUE, BLUE}} + Q_{zu}Q_{uu}^{-1}Q_{uz} \end{aligned} \tag{21}$$

Proof Since $C(\text{BLUE}, \underline{u}) = 0$ and the BLUE, like any other LUE, is a LUP, (21) follows immediately from (20). □

Here, three examples are given to see the above BLUP-relations at work.

Example 4 (The BLUPs of $\underline{y}$ and $\underline{e}$) The BLUP of $\underline{y} = Ax + \underline{e}$ is the BLUE of Ax plus the BLUP of $\underline{e}$. The BLUP of $\underline{e}$ is $\hat{\underline{e}} = Q_{yu}Q_{uu}^{-1}\underline{u}$. Substitution of $\underline{u} = B^T\underline{y}$ gives

$$\hat{\underline{e}} = Q_{yy}B(B^TQ_{yy}B)^{-1}B^T\underline{y} = \underline{y} - \hat{\underline{y}} \tag{22}$$

If we add the BLUE of $\mathsf{E}(\underline{y}) = Ax$, which is $\hat{\underline{y}} = A\hat{\underline{x}}$, we get the BLUP of $\underline{y}$ as $\underline{y} = A\hat{\underline{x}} + \hat{\underline{e}}$. Hence the BLUP of $\underline{y}$ is $\underline{y}$ itself. This is the *reproducing property* of best prediction, i.e. the best prediction of an observable is the observable itself.

Example 5 (Equality of BLUP and BLUE) We now consider $\underline{y} = A\underline{x} + \underline{n}$ in which $\underline{x}$ and $\underline{n}$ are random vectors having means $\mathsf{E}(\underline{x}) = x$ and $\mathsf{E}(\underline{n}) = 0$, respectively. The mean of $\underline{x}$ is assumed unknown. To bring the observation equations into the standard form of (1), we write

$$\underline{y} = Ax + \underline{e}\ \text{, with}\ \underline{e} = A(\underline{x} - x) + \underline{n} \tag{23}$$

Note that the misclosure vector does not depend on $\underline{x}$, but only on the measurement noise vector $\underline{n}$, i.e. $\underline{u} = B^T\underline{y} = B^T\underline{n}$. Let $\hat{\underline{x}} = A^+\underline{y}$ be the BLUE of x. Then according to (21), the BLUP of $\underline{x}$ and its error variance matrix are given as

$$\begin{aligned} \text{BLUP} &= \hat{\underline{x}} + Q_{xu}Q_{uu}^{-1}\underline{u} \\ Q_{\varepsilon_{\text{BLUP}}\varepsilon_{\text{BLUP}}} &= Q_{x-\hat{x},x-\hat{x}} - Q_{xu}Q_{uu}^{-1}Q_{ux} \end{aligned} \tag{24}$$

This shows, since $Q_{xu} = Q_{xn}B$, that the BLUP of $\underline{x}$ becomes identical to the BLUE of x if $\mathsf{C}(\underline{x}, \underline{n}) = 0$, i.e. when $\underline{x}$ and $\underline{n}$ are uncorrelated. The error variance matrix reduces then to

$$Q_{\varepsilon_{\text{BLUP}}\varepsilon_{\text{BLUP}}} = Q_{x-\hat{x},x-\hat{x}} = Q_{\hat{x}\hat{x}} - Q_{xx} \tag{25}$$

since $\mathsf{C}(\hat{\underline{x}}, \underline{x}) = \mathsf{C}(A^+\underline{y}, \underline{x}) = \mathsf{C}(\underline{x} + A^+\underline{n}, \underline{x}) = Q_{xx}$. The fact that BLUP = BLUE in this case does not mean that the two have the same quality. The BLUP should be judged as a predictor through its error variance matrix (25), whereas the BLUE should be judged as an estimator through its variance matrix $Q_{\hat{x}\hat{x}}$. We come back to the BLUE–BLUP relation when we consider their recursive forms in the next sections.

Example 6 (BLUP and BLP compared) Using BLUE $= A_z\hat{\underline{x}}$ and $Q_{zu}Q_{uu}^{-1}\underline{u} = Q_{zy}B(B^TQ_{yy}B)^{-1}B^T\underline{y} = Q_{zy}Q_{yy}^{-1}(\underline{y} - \hat{\underline{y}})$, we may write the first equation of (21) as

$$\text{BLUP} = A_z\hat{\underline{x}} + Q_{zy}Q_{yy}^{-1}(\underline{y} - \hat{\underline{y}}) \tag{26}$$

This expression makes an easy comparison with the BLP in (8) possible. It shows that the BLUP can be obtained from the BLP expression by replacing the means $\bar{z}$ and $\bar{y}$ by their BLUEs. The price to pay for such replacement lies in the larger mean squared errors. The prediction error variance matrices of the BLP and the BLUP are namely given as

$$\begin{aligned} \mathsf{C}(\underline{\varepsilon}_{\text{BLP}}, \underline{\varepsilon}_{\text{BLP}}) &= Q_{zz} - Q_{zy}Q_{yy}^{-1}Q_{yz} \\ \mathsf{C}(\underline{\varepsilon}_{\text{BLUP}}, \underline{\varepsilon}_{\text{BLUP}}) &= \mathsf{C}(\underline{\varepsilon}_{\text{BLP}}, \underline{\varepsilon}_{\text{BLP}}) + A_{z|y}Q_{\hat{x}\hat{x}}A_{z|y}^T \end{aligned} \tag{27}$$

with $A_{z|y} = A_z - Q_{zy}Q_{yy}^{-1}A$. Another consequence of estimating the unknown means in case of the BLUP is that

$$\mathsf{C}(\underline{\varepsilon}_{\text{BLUP}}, \underline{y}) \neq 0,\ \text{but}\ \mathsf{C}(\underline{\varepsilon}_{\text{BLUP}}, \underline{u}) = 0 \tag{28}$$

while $\mathsf{C}(\underline{\varepsilon}_{\text{BLP}}, \underline{y}) = 0$.

For a quick reference, a summary of the estimation–prediction relations is given in Table 1.

3 BLUE and BLUP measurement update equations

We now generalize the single observational vector $\underline{y}$ to a time series of vectorial observables, $\underline{y}_0, \ldots, \underline{y}_t$, of which the means are assumed to be linearly related to the mean of $\underline{z}$. The index refers to the time instant the data are collected.

The data are collected with the purpose of predicting $\underline{z}$ and estimating $\bar{z} = \mathsf{E}(\underline{z})$. To show how estimation and prediction are affected by the inclusion of a new observation vector $\underline{y}_i$, we present the BLUE and BLUP measurement update equations. No assumptions are yet made on the dispersion of the observables. From now on we denote a BLUE with the $\hat{.}$-symbol and a BLUP with the $\check{.}$-symbol. To show on which set of observables estimation and prediction are based, we use the notations $\hat{\underline{z}}_{|[i-1]}$ and $\check{\underline{z}}_{|[i-1]}$ when based on $\underline{y}_{[i-1]} = [\underline{y}_0^T, \ldots, \underline{y}_{i-1}^T]^T$.

Table 1 Estimation and prediction compared for linear models

$\underline{\hat{x}} = A^{+}\underline{y}$, $\underline{u} = B^{T}\underline{y}$	Estimation	Prediction
Class representation	$\text{LUE} = A_z\underline{\hat{x}} + J\underline{u}$ for some J	$\text{LUP} = A_z\underline{\hat{x}} + H\underline{u}$ for some H
Best member	$BLUE = A_z\underline{\hat{x}}$	$\text{BLUP} = A_z\underline{\hat{x}} + Q_{zu}Q_{uu}^{-1}\underline{u}$
BLUE–LUE (BLUP–LUP) formula	$\text{BLUE}=\text{LUE}-Q_{\text{LUE},u}Q_{uu}^{-1}\underline{u}$	$\text{BLUP} = \text{LUP} + Q_{z-\text{LUP},u}Q_{uu}^{-1}\underline{u}$
BLUE (BLUP) variance matrix	$Q_{\text{BLUE, BLUE}} = Q_{\text{LUE, LUE}} - Q_{\text{LUE},u}Q_{uu}^{-1}Q_{u,\text{LUE}}$	$Q_{\text{BLUP, BLUP}} = Q_{\text{LUP, LUP}} - Q_{\text{LUP},u}Q_{uu}^{-1}Q_{u,\text{LUP}} + Q_{z,u}Q_{uu}^{-1}Q_{u,z}$
Error variance matrix	$Q_{\varepsilon_{\text{BLUE}}\varepsilon_{\text{BLUE}}} = Q_{\text{LUE, LUE}} - Q_{\text{LUE},u}Q_{uu}^{-1}Q_{u,\text{LUE}}$	$Q_{\varepsilon_{\text{BLUP}}\varepsilon_{\text{BLUP}}} = Q_{z-\text{LUP},z-\text{LUP}} - Q_{z-\text{LUP},u}Q_{uu}^{-1}Q_{u,z-\text{LUP}}$
Error covariance matrices	$C(\underline{\varepsilon}_{\text{BLUE}}, \underline{u})=0,\ C(\underline{\varepsilon}_{\text{BLUE}}, \text{BLUE})=C(\underline{\varepsilon}_{\text{BLUE}}, \text{BLUP})$	$C(\underline{\varepsilon}_{\text{BLUP}}, \underline{u})=0,\ C(\underline{\varepsilon}_{\text{BLUP}}, \text{BLUE}) = C(\underline{\varepsilon}_{\text{BLUE}}, \text{BLUP})$

3.1 Uncorrelated misclosures and the statistics of the block-triangular decomposition

In the results up to now, we have emphasized the role played by the vector of misclosures $\underline{u}$, both in estimation and in prediction. We also pointed out that the vector of misclosures is not unique. Any one-to-one transformation of $\underline{u}$ produces again a vector of misclosures. Despite this nonuniqueness, the BLUE and BLUP are unique, i.e. they are invariant for any regular transformation of the vector of misclosures. This is illustrated by the identity

$$Q_{zu}Q_{uu}^{-1}\underline{u} = Q_{zv}Q_{vv}^{-1}\underline{v} \tag{29}$$

which holds for any $\underline{v} = L\underline{u}$ with invertible L.

The freedom we have in choosing the vector of misclosures makes it possible to choose a vector of misclosures with (block)diagonal variance matrix, e.g. such that they become uncorrelated from epoch to epoch. This is attractive as it generally simplifies computations. In case Q_{vv} is (block)diagonal, the multi-epoch inversion can be achieved through an epoch-by-epoch inversion of lower dimensioned matrices,

$$Q_{zv}Q_{vv}^{-1}\underline{v} = \sum_{j=1}^{t} Q_{zv_j}Q_{v_jv_j}^{-1}\underline{v}_j \tag{30}$$

Although there are different ways of making a variance matrix (block)diagonal, it is the triangular decomposition that is particularly suited for the sequential treatment of the measurement update equations. The following lemma provides the statistical proporties of the transformed misclosures that correspond to the triangular decomposition of the original misclosures' variance matrix.

Lemma 4 (Uncorrelated misclosures) *Consider the partitioned misclosure vector* $\underline{u}_{[i]} = [\underline{u}_{[i-1]}^T, \underline{u}_i^T]^T$, $i = 1, 2, \ldots$, *with* $\underline{u}_{[0]} = \underline{u}_0$, *and define* $\underline{v}_{[i]} = [\underline{v}_{[i-1]}^T, \underline{v}_i^T]^T$, $i = 1, 2, \ldots$ *as*

$$\begin{bmatrix} \underline{v}_{[i-1]} \\ \underline{v}_i \end{bmatrix} = \begin{bmatrix} L_{[i-1]} & 0 \\ -Q_{u_iu_{[i-1]}}Q_{u_{[i-1]}u_{[i-1]}}^{-1} & I \end{bmatrix} \begin{bmatrix} \underline{u}_{[i-1]} \\ \underline{u}_i \end{bmatrix} \tag{31}$$

with $L_{[0]} = I$. *Then*

(a) $\underline{v}_i = \underline{u}_i - \underline{\check{u}}_{i|[i-1]}$ *is the sequential predictor error of* $\underline{u}_i$
(b) $\underline{v}_{[i]} = L_{[i]}\underline{u}_{[i]}$, *with* $L_{[i]}$ *unit lower (block)triangular*
(c) $Q_{v_{[i]}v_{[i]}} = \text{blockdiag}(Q_{v_0v_0}, \ldots, Q_{v_iv_i})$

Proof (*a*) The BLUP of $\underline{u}_i$ based on the previous misclosure vector $\underline{u}_{[i-1]}$ is $\underline{\check{u}}_{i|[i-1]} = Q_{u_iu_{[i-1]}}Q_{u_{[i-1]}u_{[i-1]}}^{-1}\underline{u}_{[i-1]}$. Hence, $\underline{v}_i = \underline{u}_i - \underline{\check{u}}_{i|[i-1]}$ is the corresponding prediction error. (*b*) The transformation matrix $L_{[i]}$ of (31) is lower triangular with 1's on the diagonal. (*c*) We have $\mathsf{C}(\underline{v}_i, \underline{u}_{[i-1]}) = \mathsf{C}(\underline{u}_i - \underline{\check{u}}_{i|[i-1]}, \underline{u}_{[i-1]}) = 0$, since the BLUP prediction error is uncorrelated with its misclosure vector. Substitution of $\underline{u}_{[i-1]} = L_{[i-1]}^{-1}\underline{v}_{[i-1]}$ gives $\mathsf{C}(\underline{v}_i, \underline{v}_{[i-1]}) = 0$. Hence, the variance matrix of $\underline{v}_{[i]}$ is block diagonal. □

As the above result shows, the random vectors $\underline{v}_i$, $i = 0, 1, \ldots$, are uncorrelated misclosures and at the same time sequential prediction errors of their correlated counterparts.

3.2 Measurement update equations

We are now in a position to use the misclosure vector $\underline{v}_{[i]} = [\underline{v}_{[i-1]}^T, \underline{v}_i^T]^T$ and the blockdiagonal structure of its variance matrix, $Q_{v_{[i]}v_{[i]}} = \text{blockdiag}(Q_{v_{[i-1]}v_{[i-1]}}, Q_{v_iv_i})$, to formulate the BLUE and BLUP measurement update equations.

Lemma 5 (BLUE–BLUP measurement update) *Given the new observation vector* $\underline{y}_i$, *the BLUE* $\underline{\hat{z}}_{|[i-1]}$ *and BLUP* $\underline{\check{z}}_{|[i-1]}$ *of* $\bar{z}$ *and* $\underline{z}$, *respectively, are updated as*

$$\begin{aligned} \underline{\hat{z}}_{|[i]} &= \underline{\hat{z}}_{|[i-1]} - Q_{\hat{z}_{|[i-1]}v_i}Q_{v_iv_i}^{-1}\underline{v}_i, \\ Q_{\hat{z}_{|[i]}\hat{z}_{|[i]}} &= Q_{\hat{z}_{|[i-1]}\hat{z}_{|[i-1]}} - Q_{\hat{z}_{|[i-1]}v_i}Q_{v_iv_i}^{-1}Q_{v_i\hat{z}_{|[i-1]}} \end{aligned} \tag{32}$$

and

$$\begin{aligned} \underline{\check{z}}_{|[i]} &= \underline{\check{z}}_{|[i-1]} + Q_{z-\check{z}_{|[i-1]},v_i}Q_{v_iv_i}^{-1}\underline{v}_i, \\ P_{\check{z}_{|[i]}\check{z}_{|[i]}} &= P_{\check{z}_{|[i-1]}\check{z}_{|[i-1]}} - Q_{z-\check{z}_{|[i-1]},v_i}Q_{v_iv_i}^{-1}Q_{v_i,z-\check{z}_{|[i-1]}} \end{aligned} \tag{33}$$

Proof Since the BLUE $\underline{\hat{z}}_{|[i-1]}$ is a LUE where it is based on $\underline{y}_{[i]}$, we can apply the BLUE–LUE formula of Corollary 1, and obtain

$$\hat{\underline{z}}_{|[i]} = \hat{\underline{z}}_{|[i-1]} - Q_{\hat{z}_{|[i-1]}v_{[i]}} Q^{-1}_{v_{[i]}v_{[i]}} \underline{v}_{[i]},$$
$$Q_{\hat{z}_{|[i]}\hat{z}_{|[i]}} = Q_{\hat{z}_{|[i-1]}\hat{z}_{|[i-1]}} - Q_{\hat{z}_{|[i-1]}v_{[i]}} Q^{-1}_{v_{[i]}v_{[i]}} Q_{v_{[i]}\hat{z}_{|[i-1]}} \quad (34)$$

With $\mathsf{C}(\hat{\underline{z}}_{|[i-1]}, \underline{v}_{[i-1]}) = 0$ and the blockdiagonality of $Q_{v_{[i]}v_{[i]}}$, the above equation simplifies to Eq. (32). The proof for the predictor goes likewise. □

As the lemma shows, the contribution to the difference between two succeeding estimators and their respective predictors, $\hat{\underline{z}}_{|[i]}$ and $\hat{\underline{z}}_{|[i-1]}$ or $\check{\underline{z}}_{|[i]}$ and $\check{\underline{z}}_{|[i-1]}$, is provided by $\underline{v}_i$. Since $\underline{v}_i$ is uncorrelated with $\underline{v}_{[i-1]} = L_{[i-1]}\underline{u}_{[i-1]}$, it cannot be predicted by the previous information, and therefore, it contains truly *new* information. The term *innovation* for $\underline{v}_i$ was first independently introduced, in case of stationary time series, by Bode and Shannon (1950) and Zadeh and Ragazzini (1950), and later further used for known-mean, non-stationary time series by Kailath (1968).

3.3 Partitioned linear model and predicted residuals

Note that the results of the above two Lemmas 4 and 5 do not depend on the linear model structure of the observables. We now introduce this structure by means of the partitioned linear model

$$\begin{bmatrix} \underline{y}_{[i-1]} \\ \underline{y}_i \end{bmatrix} = \begin{bmatrix} A_{[i-1]} \\ A_i \end{bmatrix} x + \begin{bmatrix} \underline{e}_{[i-1]} \\ \underline{e}_i \end{bmatrix} \quad (35)$$

with mean and dispersion

$$\mathsf{E}\left(\begin{bmatrix} \underline{e}_{[i-1]} \\ \underline{e}_i \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathsf{D}\left(\begin{bmatrix} \underline{y}_{[i-1]} \\ \underline{y}_i \end{bmatrix}\right) = \begin{bmatrix} Q_{y_{[i-1]}y_{[i-1]}} & Q_{y_{[i-1]}y_i} \\ Q_{y_i y_{[i-1]}} & Q_{y_i y_i} \end{bmatrix}$$

We make use of this structure to get a further interpretation of the innovation vector $\underline{v}_i$.

Lemma 6 (Estimation and prediction residuals) *Let the transpose of the basis matrix $B_{[i]}$ of the null space of $A^T_{[i]} = [A^T_{[i-1]}, A^T_i]$ be chosen as*

$$B^T_{[i]} = \begin{bmatrix} B^T_{[i-1]} & 0 \\ -A_i A^+_{[i-1]} & I \end{bmatrix} \quad (36)$$

with $A^+_{[i-1]} = \left(A^T_{[i-1]} Q^{-1}_{y_{[i-1]}y_{[i-1]}} A_{[i-1]}\right)^{-1} A^T_{[i-1]} Q^{-1}_{y_{[i-1]}y_{[i-1]}}$. *Then*

$$\begin{aligned} (a)\ & \underline{u}_i = \underline{y}_i - \hat{\underline{y}}_{i|[i-1]} \\ (b)\ & \underline{v}_i = \underline{y}_i - \check{\underline{y}}_{i|[i-1]} \end{aligned} \quad (37)$$

with BLUE $\hat{\underline{y}}_{i|[i-1]} = A_i \hat{\underline{x}}_{|[i-1]}$ *and BLUP* $\check{\underline{y}}_{i|[i-1]} = \hat{\underline{y}}_{i|[i-1]} + Q_{y_i u_{[i-1]}} Q^{-1}_{u_{[i-1]}u_{[i-1]}} \underline{u}_{[i-1]}$.

Proof (*a*) With (36), it follows from $\underline{u}_{[i]} = B^T_{[i]} \underline{y}_{[i]} = [\underline{u}^T_{[i-1]}, \underline{u}^T_i]^T$ that $\underline{u}_i = \underline{y}_i - A_i A^+_{[i-1]} y_{[i-1]}$. Hence, with $\hat{\underline{y}}_{i|[i-1]} = A_i \hat{\underline{x}}_{|[i-1]}$ and $\hat{\underline{x}}_{|[i-1]} = A^+_{[i-1]} y_{[i-1]}$ the result follows. (*b*) Since $\mathsf{C}(\hat{\underline{x}}_{|[i-1]}, \underline{u}_{[i-1]}) = 0$, we have $\mathsf{C}(\underline{u}_i, \underline{u}_{[i-1]}) = \mathsf{C}(\underline{y}_i - A_i \hat{\underline{x}}_{|[i-1]}, \underline{u}_{[i-1]}) = \mathsf{C}(\underline{y}_i, \underline{u}_{[i-1]})$ and thus $Q_{u_i, u_{[i-1]}} = Q_{y_i, u_{[i-1]}}$. We may therefore write $\check{\underline{u}}_{i|[i-1]} = Q_{y_i u_{[i-1]}} Q^{-1}_{u_{[i-1]}u_{[i-1]}} \underline{u}_{[i-1]}$. Substitution into $\underline{v}_i = \underline{u}_i - \check{\underline{u}}_{i|[i-1]}$ gives, with $\underline{u}_i = \underline{y}_i - \hat{\underline{y}}_{i|[i-1]}$ and $\check{\underline{y}}_{i|[i-1]} = \hat{\underline{y}}_{i|[i-1]} + Q_{y_i u_{[i-1]}} Q^{-1}_{u_{[i-1]}u_{[i-1]}} \underline{u}_{[i-1]}$, the stated result. □

The above result (37) shows that the innovation vector $\underline{v}_i$ is not only the sequential prediction error of $\underline{u}_i$ (cf. Lemma 4), but also, in case the basis matrix is chosen as (36), that of the observation vector $\underline{y}_i$. The innovation vector $\underline{v}_i$ will therefore from now on be referred to as the *predicted residual* of $\underline{y}_i$.

Note, since $Q_{y_i u_{[i-1]}} = 0$ if $\mathsf{C}(\underline{y}_i, \underline{y}_j) = 0,\ i \neq j$, that the estimator and predictor coincide, $\hat{\underline{y}}_{i|[i-1]} = \check{\underline{y}}_{i|[i-1]}$, if the observables are uncorrelated. In that case, we have $\underline{u}_i = \underline{v}_i = \underline{y}_i - A_i \hat{\underline{x}}_{|[i-1]}$.

Example 7 (Recursive least-squares estimation) If we assume uncorrelated observation vectors $\underline{y}_i$, i.e. $\mathsf{C}(\underline{y}_i, \underline{y}_j) = 0,\ i \neq j$, and take $A_z = I$, then (32) reduces, with $\underline{v}_i = \underline{y}_i - A_i \hat{\underline{x}}_{|[i-1]}$ and $Q_{\hat{x}_{|[i-1]}v_i} = -Q_{\hat{x}_{|[i-1]}\hat{x}_{|[i-1]}} A^T_i$, to the well-known recursive least-squares solution

$$\hat{\underline{x}}_{|[i]} = \hat{\underline{x}}_{|[i-1]} + Q_{\hat{x}_{|[i-1]}\hat{x}_{|[i-1]}} A^T_i Q^{-1}_{v_i v_i} (\underline{y}_i - A_i \hat{\underline{x}}_{|[i-1]}) \quad (38)$$

4 Recursive BLUP and the Kalman filter

We now consider as partitioned model the measurement and dynamic models that form the basis of the well-known recursive Kalman filter. However, instead of the standard assumption that the means of the random state-vectors are known, we assume them to be unknown. Since the model can be brought into the linear model form, the BLUE and BLUP results of the previous sections directly apply. The recursive BLUP is shown to follow the Kalman filter updates, albeit with an initialization that is different from that of the known-mean, standard Kalman filter.

4.1 Model assumptions

First, we state the assumptions concerning the measurement and dynamic models.

The measurement model The link between the random vector of observables $\underline{y}_i$ and the random state-vector $\underline{x}_i$ is assumed given as

$$\underline{y}_i = A_i \underline{x}_i + \underline{n}_i,\ i = 0, 1, \ldots, t, \quad (39)$$

together with

$$\mathsf{E}(\underline{x}_0) = x_0 \text{ (unknown)}, \quad \mathsf{E}(\underline{n}_i) = 0, \quad (40)$$

and

$$\mathsf{C}(\underline{x}_0, \underline{n}_i) = 0, \quad \mathsf{C}(\underline{n}_i, \underline{n}_j) = R_i \delta_{i,j},\ i = 0, 1, \ldots, t \quad (41)$$

with $\delta_{i,j}$ being the Kronecker delta. Thus, the zero-mean measurement noise $\underline{n}_i$ is assumed to be uncorrelated in time and to be uncorrelated with the initial state-vector $\underline{x}_0$.

The dynamic model The linear dynamic model, describing the time evolution of the random state-vector $\underline{x}_i$, is given as

$$\underline{x}_i = \Phi_{i,i-1}\underline{x}_{i-1} + \underline{d}_i,\ i = 1, 2, \ldots, t \tag{42}$$

with

$$\mathsf{E}(\underline{d}_i) = 0,\ \mathsf{C}(\underline{x}_0, \underline{d}_i) = 0, \tag{43}$$

$$\mathsf{C}(\underline{d}_i, \underline{n}_j) = 0,\ \mathsf{C}(\underline{d}_i, \underline{d}_j) = S_i\delta_{i,j},\ i, j = 1, 2, \ldots, t \tag{44}$$

where $\Phi_{i,i-1}$ denotes the transition matrix and the random vector $\underline{d}_i$ is the system noise. The system noise $\underline{d}_i$ is thus also assumed to have a zero mean, to be uncorrelated in time and to be uncorrelated with the initial state-vector and the measurement noise.

The above model equations are formulated on an epoch-by-epoch basis (cf. 39 and 42). To establish the link with the linear model formulation as used in the previous sections, one can obtain the corresponding multi-epoch linear model of (39) and (42) by defining the observation vector as $\underline{y} = [\underline{y}_0^T,\ \underline{y}_1^T,\ \ldots,\ \underline{y}_t^T]^T$. This is shown in Table 2. Hence, we can now directly apply the BLUE–BLUP results of the previous sections for predicting the random state-vector $\underline{x}_t$ and estimating its unknown mean $\mathsf{E}(\underline{x}_t) = x_t$. From now on we denote the variance matrix of the BLUE $\hat{\underline{x}}_{t|[t]}$ with $Q_{t|[t]}$, whereas the *error* variance matrix of the BLUP $\check{\underline{x}}_{t|[t]}$ is denoted by $P_{t|[t]}$. Similar notation is employed for $Q_{t|[t-1]}$ and $P_{t|[t-1]}$. We start with the initialization, i.e. the case $t = 0$, where we assume that the data vector $\underline{y}_0$ contains the complete information content to determine the unknown mean x_0, i.e. A_0 is of full column rank.

4.2 Initialization

For $t = 0$, we have the linear model $\underline{y}_0 = A_0\underline{x}_0 + \underline{n}_0$, which may also be written as $\underline{y}_0 = A_0x_0 + \underline{e}_0$, with $\underline{e}_0 = A_0(\underline{x}_0 - x_0) + \underline{n}_0$. In general the BLUE of x_0 differs from the BLUP of $\underline{x}_0$. In our case, however, the two coincide. As shown in Example 5, the predictor and estimator, and their (error) variance matrices, are simply related as

$$\check{\underline{x}}_{0|0} = \hat{\underline{x}}_{0|0} \quad \text{and} \quad P_{0|0} = Q_{0|0} - Q_{x_0x_0} \tag{45}$$

when $\underline{x}_0$ and $\underline{n}_0$ are assumed uncorrelated. In the next lemma, these expressions are further worked out in A_0, R_0 and $Q_{x_0x_0}$.

Lemma 7 (BLUE–BLUP initialization) *Let the linear model at time $t = 0$ be given as $\underline{y}_0 = A_0\underline{x}_0 + \underline{n}_0$, with unknown mean state-vector $\mathsf{E}(\underline{x}_0) = x_0$, zero-mean noise vector $\mathsf{E}(\underline{n}_0) = 0$ and variance matrix $Q_{y_0y_0} = A_0Q_{x_0x_0}A_0^T + R_0$.*

Table 2 Linear model formulation for estimation and prediction

Linear model		Linear model form of (39) and (42)
Model eqs.	$\begin{bmatrix} \underline{y} \\ \underline{z} \end{bmatrix} = \begin{bmatrix} A \\ A_z \end{bmatrix} x + \begin{bmatrix} \underline{e} \\ \underline{e}_z \end{bmatrix}$	$\underline{y} = \left[\underline{y}_0^T,\ \underline{y}_1^T,\ \ldots,\ \underline{y}_t^T\right]^T,\ A = \left[A_0^T,\ (A_1\Phi_{1,0})^T,\ \ldots,\ (A_t\Phi_{t,0})^T\right]^T,$ $\underline{e} = \left[[A_0(\underline{x}_0 - x_0) + \underline{n}_0]^T, [A_1\{\Phi_{1,0}(\underline{x}_0 - x_0) + \underline{d}_1\} + \underline{n}_1]^T, \ldots, [A_t\{\Phi_{t,0}(\underline{x}_0 - x_0) + \sum_{i=1}^t \Phi_{t,i}\,\underline{d}_i\} + \underline{n}_t]^T\right]^T$ $\underline{z} = \underline{x}_t,\ \ A_z = \Phi_{t,0},\ \ \underline{e}_z = \Phi_{t,0}(\underline{x}_0 - x_0) + \sum_{i=1}^t \Phi_{t,i}\,\underline{d}_i$
Mean, Dispersion	$\mathsf{E}\left(\begin{bmatrix} \underline{e} \\ \underline{e}_z \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathsf{D}\left(\begin{bmatrix} \underline{y} \\ \underline{z} \end{bmatrix}\right) = \begin{bmatrix} Q_{ee} & Q_{ee_z} \\ Q_{e_ze} & Q_{e_ze_z} \end{bmatrix}$	$Q_{ee} = \begin{bmatrix} A_0Q_{x_0x_0}A_0^T + R_0 & & & \\ A_1\Phi_{1,0}Q_{x_0x_0}A_0^T & A_1(\Phi_{1,0}Q_{x_0x_0}\Phi_{1,0}^T + S_1)A_1^T + R_1 & & \\ \vdots & \vdots & \ddots & \\ A_t\Phi_{t,0}Q_{x_0x_0}A_0^T & A_t(\Phi_{t,0}Q_{x_0x_0}\Phi_{1,0}^T + \Phi_{t,1}S_1)A_1^T & \ldots & A_t(\Phi_{t,0}Q_{x_0x_0}\Phi_{t,0}^T + \sum_{i=1}^t \Phi_{t,i}S_i\Phi_{t,i}^T)A_t^T + R_t \end{bmatrix}$ $Q_{e_ze} = \left[\Phi_{t,0}Q_{x_0x_0}A_0^T, (\Phi_{t,0}Q_{x_0x_0}\Phi_{1,0}^T + \Phi_{t,1}S_1)A_1^T, \ldots, (\Phi_{t,0}Q_{x_0x_0}\Phi_{t,0}^T + \sum_{i=1}^t \Phi_{t,i}S_i\Phi_{t,i}^T)A_t^T\right]$ $Q_{e_ze_z} = \Phi_{t,0}Q_{x_0x_0}\Phi_{t,0}^T + \sum_{i=1}^t \Phi_{t,i}S_i\Phi_{t,i}^T$
BLUE of $\bar{z}$	$\hat{\underline{z}} = A_z\hat{\underline{x}} = A_z(A^TQ_{yy}^{-1}A)^{-1}A^TQ_{yy}^{-1}\underline{y}$	Recursive BLUE of x_t : $\hat{\underline{x}}_{t\|t-1},\ \hat{\underline{x}}_{t\|t}$ (cf. 60, 61, 62)
BLUP of $\underline{z}$	$\check{\underline{z}} = A_z\hat{\underline{x}} + Q_{zy}Q_{yy}^{-1}(\underline{y} - A\hat{\underline{x}})$	Recursive BLUP of $\underline{x}_t$: $\check{\underline{x}}_{t\|t-1},\ \check{\underline{x}}_{t\|t}$ (cf. 50, 51, 52)

Then, the BLUE of x_0 *is equal to the BLUP of* $\underline{x}_0$ *and given as*

$$\hat{x}_{0|0} = \check{x}_{0|0} = (A_0^T R_0^{-1} A_0)^{-1} A_0^T R_0^{-1} \underline{y}_0 \tag{46}$$

with (error)variance matrices

$$Q_{0|0} = (A^T R_0^{-1} A_0)^{-1} + Q_{x_0x_0}$$
$$P_{0|0} = (A_0^T R_0^{-1} A_0)^{-1} \tag{47}$$

Proof The equality of the BLUE and BLUP in (46) is due to the zero correlation between state-vector and measurement noise vector. That R_0^{-1}, instead of $Q_{y_0y_0}^{-1}$, may be used as weight matrix in the least-squares formula of (46) follows from the matrix identity $(A_0^T Q_{y_0y_0}^{-1} A_0)^{-1} A_0^T Q_{y_0y_0}^{-1} \underline{y}_0 = (A_0^T R_0^{-1} A_0)^{-1} A_0^T R_0^{-1} \underline{y}_0$ for $Q_{y_0y_0} = A_0 Q_{x_0x_0} A_0^T + R_0$. Application of the (co)variance propagation law to (46) gives the (error)variance matrices of (47). □

As the above result shows, the BLUP of $\underline{x}_0$ is independent of its variance matrix $Q_{x_0x_0}$. This variance matrix is therefore not needed for the initialization. It is not needed for computing $\check{\underline{x}}_{0|0}$, nor for its error variance matrix $P_{0|0}$. This is a marked difference with the standard formulation of the Kalman filter where the mean of the state-vector is assumed known. In that case, prediction is based on the BLP and the initialization takes the form

$$\check{x}_{0|0}^{\text{BLP}} = \mathsf{E}(\underline{x}_0), \quad P_{0|0}^{\text{BLP}} = \mathsf{C}(\underline{x}_0 - \check{x}_{0|0}^{\text{BLP}}, \underline{x}_0 - \check{x}_{0|0}^{\text{BLP}}) = Q_{x_0x_0} \tag{48}$$

Hence, in that case the known mean is taken as the initial prediction of the random state-vector. As a consequence, the variance matrix $Q_{x_0x_0}$ is then needed as it equals the error variance matrix.

Although $Q_{x_0x_0}$ is not needed for the BLUP, the above lemma shows that it is needed for the BLUE. Not so much for computing the BLUE, but for describing its quality by means of its variance matrix $Q_{0|0}$.

4.3 Recursive BLUP

Recursion of the BLUP $\check{\underline{x}}_{t|[t]}$ is possible since the predictors of the zero-mean measurement and system noise are identically zero, $\check{\underline{n}}_{t|[t-1]} = 0$ and $\check{\underline{d}}_{t|[t-1]} = 0$. This is a consequence of having measurement noise and system noise that are both uncorrelated with the observables and state-vectors of the previous epochs. They are thus also uncorrelated with the predicted residuals of these epochs.

Since $\check{\underline{d}}_{t|[t-1]} = 0$, the predictor $\check{\underline{x}}_{t|[t-1]}$ can be computed directly from $\check{\underline{x}}_{t-1|[t-1]}$, thus providing the time update. Similarly, since $\check{\underline{n}}_{t|[t-1]} = 0$, the predicted residual $\underline{v}_t = \underline{y}_t - \check{\underline{y}}_{t|[t-1]}$ can be computed directly from $\underline{y}_t$ and $\check{\underline{x}}_{t|[t-1]}$, thus providing, in combination with (33), the measurement update of $\check{\underline{x}}_{t|[t-1]}$.

Lemma 8 (Predicted residuals) *For the measurement and dynamic model (39) and (42), the predicted residual vector and its variance matrix are given as*

$$\underline{v}_t = \underline{y}_t - A_t \check{\underline{x}}_{t|[t-1]}$$
$$Q_{v_tv_t} = R_t + A_t P_{t|[t-1]} A_t^T \tag{49}$$

Proof As the BLUP of a linear function is the linear function of the BLUP, the BLUP of $\underline{y}_t = A_t \underline{x}_t + \underline{n}_t$ is $\check{\underline{y}}_{t|[t-1]} = A_t \check{\underline{x}}_{t|[t-1]} + \check{\underline{n}}_{t|[t-1]}$, with $\check{\underline{n}}_{t|[t-1]} = 0$, since $\underline{n}_t$ is zero mean and uncorrelated with the previous predicted residuals. Substitution of $\check{\underline{y}}_{t|[t-1]} = A_t \check{\underline{x}}_{t|[t-1]}$ into $\underline{v}_t = \underline{y}_t - \check{\underline{y}}_{t|[t-1]}$ proves the first equation of (49). The second equation follows from an application of the variance propagation law to $\underline{v}_t = A_t(\underline{x}_t - \check{\underline{x}}_{t|[t-1]}) + \underline{n}_t$. □

The steps for the recursion of $\check{\underline{x}}_{t|t}$ can now be summarized as follows.

Theorem 2 (a) (Recursive BLUP) *The three steps of the recursive state-vector prediction are given as*

$$\textit{Initialization:}\quad \begin{aligned} \check{\underline{x}}_{0|0} &= (A_0^T R_0^{-1} A_0)^{-1} A_0^T R_0^{-1} \underline{y}_0, \\ P_{0|0} &= (A_0^T R_0^{-1} A_0)^{-1} \end{aligned} \tag{50}$$

$$\textit{Time update:}\quad \begin{aligned} \check{\underline{x}}_{t|[t-1]} &= \Phi_{t,t-1} \check{\underline{x}}_{t-1|[t-1]}, \\ P_{t|[t-1]} &= \Phi_{t,t-1} P_{t-1|[t-1]} \Phi_{t,t-1}^T + S_t \end{aligned} \tag{51}$$

$$\textit{Measurement update:}\quad \begin{aligned} \check{\underline{x}}_{t|[t]} &= \check{\underline{x}}_{t|[t-1]} + K_t \underline{v}_t, \\ P_{t|[t]} &= P_{t|[t-1]} - K_t Q_{v_tv_t} K_t^T \end{aligned} \tag{52}$$

with gain matrix $K_t = P_{t|[t-1]} A_t^T Q_{v_tv_t}^{-1}$.

Proof The *initialization* was already proven in (46) and (47). To find the *time update*, we determine the BLUP of $\underline{x}_t = \Phi_{t,t-1}\underline{x}_{t-1} + \underline{d}_t$ as $\check{\underline{x}}_{t|[t-1]} = \Phi_{t,t-1}\check{\underline{x}}_{t-1|[t-1]} + \check{\underline{d}}_{t|[t-1]}$, with $\check{\underline{d}}_{t|[t-1]} = 0$, since $\underline{d}_t$ is zero mean and uncorrelated with the previous predicted residuals. This proves the first equation of (51). The second expression follows by applying the variance propagation law to $\underline{x}_t - \check{\underline{x}}_{t|[t-1]} = \Phi_{t,t-1}(\underline{x}_{t-1} - \check{\underline{x}}_{t-1|[t-1]}) + \underline{d}_t$ and using the relation $\mathsf{C}(\underline{x}_{t-1} - \check{\underline{x}}_{t-1|[t-1]}, \underline{d}_t) = 0$. To determine the *measurement update*, we apply (33), noting that for $\underline{z} = \underline{x}_t$, we need the covariance matrix $\mathsf{C}(\underline{x}_t - \check{\underline{x}}_{t|[t-1]}, v_t)$. With $\underline{v}_t = A_t(\underline{x}_t - \check{\underline{x}}_{t|[t-1]}) + \underline{n}_t$, this gives $\mathsf{C}(\underline{x}_t - \check{\underline{x}}_{t|[t-1]}, v_t) = P_{t|[t-1]} A_t^T$. Substitution into (33) proves (52). □

This result shows that apart from the initialization, the structure of the recursive BLUP is identical to that of the Kalman filter. Although the two expressions of the initialization (50) may suggest otherwise, it is important to note that $P_{0|0}$ is *not* the variance matrix of $\check{\underline{x}}_{0|0}$, but rather its *error* variance matrix. The variance matrix of $\underline{y}_0$ is namely not R_0, but $A_0 Q_{x_0x_0} A_0^T + R_0$. The variance matrix of $\check{\underline{x}}_{0|0}$ is therefore equal to the sum $Q_{x_0x_0} + P_{0|0}$ and not equal to $P_{0|0}$.

As already pointed out earlier, the BLUP-initialization does not require the variance matrix $Q_{x_0x_0}$ of the initial state-vector $\underline{x}_0$. In fact, as the theorem now shows, this variance matrix is not needed at all. Hence, $Q_{x_0x_0}$ can take *any* value (e.g. 0 or ∞) without it having any effect on the result and quality of the recursive BLUP. This is in marked contrast to the standard Kalman filter.

5 The BLUE–BLUP recursion

Next to the prediction, we now present the recursive BLUE solution. This extension of the 'Kalman-filter' theory is a consequence of our relaxing assumptions that the means of the random state-vectors are unknown. In the standard Kalman filter set-up with known state-vector means, this difference between estimation and prediction does not occur since one is then only left with BLP instead of with BLUP of the state-vectors.

5.1 Time evolution of the error covariances

In order to develop the recursion for the BLUE $\hat{\underline{x}}_{t|[t-1]}$, we first determine the time evolution of the BLUE–BLUP error covariance matrices

$$C_{t|[t]} = \mathsf{C}(x_t - \hat{\underline{x}}_{t|[t]}, \underline{x}_t - \check{\underline{x}}_{t|[t]}) \quad \text{and}$$
$$C_{t|[t-1]} = \mathsf{C}(x_t - \hat{\underline{x}}_{t|[t-1]}, \underline{x}_t - \check{\underline{x}}_{t|[t-1]}) \tag{53}$$

The following lemma shows how these error covariance matrices can be computed recursively.

Lemma 9 (BLUE–BLUP error covariance) *The time evolution of the error covariance matrices $C_{t|[t]}$ and $C_{t|[t-1]}$ is given as:*

$$\text{Initialization:} \quad C_{0|0} = P_{0|0} \tag{54}$$

$$\text{Time update:} \quad C_{t|[t-1]} = \Phi_{t,t-1} C_{t-1|[t-1]} \Phi_{t,t-1}^T \tag{55}$$

$$\text{Measurement update:} \quad C_{t|[t]} = C_{t|[t-1]} (I - K_t A_t)^T \tag{56}$$

Proof For the *initialization* ($t = 0$) we have $C_{0|0} = \mathsf{C}(x_0 - \hat{\underline{x}}_{0|0}, \underline{x}_0 - \check{\underline{x}}_{0|0}) = Q_{0|0} - Q_{x_0x_0} = P_{0|0}$, since $\hat{\underline{x}}_{0|0} = \check{\underline{x}}_{0|0}$ and $\mathsf{C}(\hat{\underline{x}}_{0|0}, \underline{x}_0) = \mathsf{C}(\underline{x}_0, \underline{x}_0)$. This proves (54). For the *time update* we have, with $\hat{\underline{\varepsilon}}_{t|[t-1]} = x_t - \hat{\underline{x}}_{t|[t-1]}$ and $\check{\underline{\varepsilon}}_{t|[t-1]} = \underline{x}_t - \check{\underline{x}}_{t|[t-1]}$, that

$$\begin{aligned} C_{t|[t-1]} &= \mathsf{C}(\hat{\underline{\varepsilon}}_{t|[t-1]}, \check{\underline{\varepsilon}}_{t|[t-1]}) \\ &= \mathsf{C}(\Phi_{t,t-1}\hat{\underline{\varepsilon}}_{t-1|[t-1]}, \Phi_{t,t-1}\check{\underline{\varepsilon}}_{t-1|[t-1]} + \underline{d}_t) \\ &= \Phi_{t,t-1} C_{t-1|[t-1]} \Phi_{t,t-1}^T \end{aligned} \tag{57}$$

which proves (55). To prove the *measurement update*, we make use of Corollary 1. Since $\hat{\underline{\varepsilon}}_{t|[t-1]}$ and $\check{\underline{\varepsilon}}_{t|[t-1]}$ are both uncorrelated with $\underline{v}_i$ for $i < t$, it follows from Corollary 1, with $\hat{\underline{\varepsilon}}_{t|[t]} = x_t - \hat{\underline{x}}_{t|[t]}$ and $\check{\underline{\varepsilon}}_{t|[t]} = \underline{x}_t - \check{\underline{x}}_{t|[t]}$, that

$$\begin{aligned} \mathsf{C}(\hat{\underline{\varepsilon}}_{t|[t]}, \check{\underline{\varepsilon}}_{t|[t]}) &= \mathsf{C}(\hat{\underline{\varepsilon}}_{t|[t-1]}, \check{\underline{\varepsilon}}_{t|[t-1]}) \\ &\quad - \mathsf{C}(\hat{\underline{\varepsilon}}_{t|[t-1]}, \underline{v}_t) Q_{v_tv_t}^{-1} \mathsf{C}(\check{\underline{\varepsilon}}_{t|[t-1]}, \underline{v}_t)^T \end{aligned} \tag{58}$$

Furthermore we have, with $\underline{v}_t = A_t \check{\underline{\varepsilon}}_{t|[t-1]} + \underline{n}_t$,

$$\begin{aligned} \mathsf{C}(\hat{\underline{\varepsilon}}_{t|[t-1]}, \underline{v}_t) &= \mathsf{C}(\hat{\underline{\varepsilon}}_{t|[t-1]}, \check{\underline{\varepsilon}}_{t|[t-1]}) A_t^T = C_{t|[t-1]} A_t^T \\ \mathsf{C}(\check{\underline{\varepsilon}}_{t|[t-1]}, \underline{v}_t) &= \mathsf{C}(\check{\underline{\varepsilon}}_{t|[t-1]}, \check{\underline{\varepsilon}}_{t|[t-1]}) A_t^T = P_{t|[t-1]} A_t^T \end{aligned} \tag{59}$$

Substitution into (58) gives the measurement update of the error covariance matrix as $C_{t|[t]} = C_{t|[t-1]}(I - A_t^T Q_{v_tv_t}^{-1} A_t P_{t|[t-1]})$, which proves (56). □

5.2 Recursive BLUE

With the help of the recursion of these error covariance matrices, it becomes possible to set up the recursion for the BLUE of the mean state-vectors $x_t = \mathsf{E}(\underline{x}_t)$.

Theorem 2 (b) (Recursive BLUE) *The three steps of the recursive mean state-vector estimation are given as*

$$\text{Initialization:} \quad \begin{aligned} \hat{\underline{x}}_{0|0} &= (A_0^T R_0^{-1} A_0)^{-1} A_0^T R_0^{-1} \underline{y}_0, \\ Q_{0|0} &= Q_{x_0x_0} + P_{0|0} \end{aligned} \tag{60}$$

$$\text{Time update:} \quad \begin{aligned} \hat{\underline{x}}_{t|[t-1]} &= \Phi_{t,t-1} \hat{\underline{x}}_{t-1|[t-1]}, \\ Q_{t|[t-1]} &= \Phi_{t,t-1} Q_{t-1|[t-1]} \Phi_{t,t-1}^T \end{aligned} \tag{61}$$

$$\text{Measurement update:} \quad \begin{aligned} \hat{\underline{x}}_{t|[t]} &= \hat{\underline{x}}_{t|[t-1]} + G_t \underline{v}_t, \\ Q_{t|[t]} &= Q_{t|[t-1]} - G_t Q_{v_tv_t} G_t^T \end{aligned} \tag{62}$$

with gain matrix $G_t = C_{t|[t-1]} A_t^T Q_{v_tv_t}^{-1}$.

Proof The *initialization* was already proven in (46) and (47). Since the BLUE of a linear function is the linear function of the BLUE, also the *time update* (61) directly follows. To determine the *measurement update*, we apply (32), noting that for $\bar{z} = \mathsf{E}(\underline{x}_t)$, we need the covariance matrix $\mathsf{C}(x_t - \hat{\underline{x}}_{t|[t-1]}, v_t)$. With $\underline{v}_t = A_t(\underline{x}_t - \check{\underline{x}}_{t|[t-1]}) + \underline{n}_t$, this gives $\mathsf{C}(x_t - \hat{\underline{x}}_{t|[t-1]}, v_t) = C_{t|[t-1]} A_t^T$. Substitution into (32) proves (62). □

Compare this BLUE recursion with the BLUP recursion of Theorem 2(a). Both look very similar. They have the same structure and they even have the same initialization ($\hat{\underline{x}}_{0|0} = \check{\underline{x}}_{0|0}$) and the same time update ($\hat{\underline{x}}_{t|[t-1]} = \Phi_{t,t-1} \hat{\underline{x}}_{t-1|[t-1]}$ versus $\check{\underline{x}}_{t|[t-1]} = \Phi_{t,t-1} \check{\underline{x}}_{t-1|[t-1]}$). They differ however in the variance matrices and in their measurement updates. In case of the BLUP, the error variance matrix $P_{t|[t-1]}$ is used both in the computation of the gain K_t and in the quality evaluation of the predictor. In case of the BLUE, however, the quality of the estimator $\hat{\underline{x}}_{t|[t-1]}$ is described by $Q_{t|[t-1]}$,

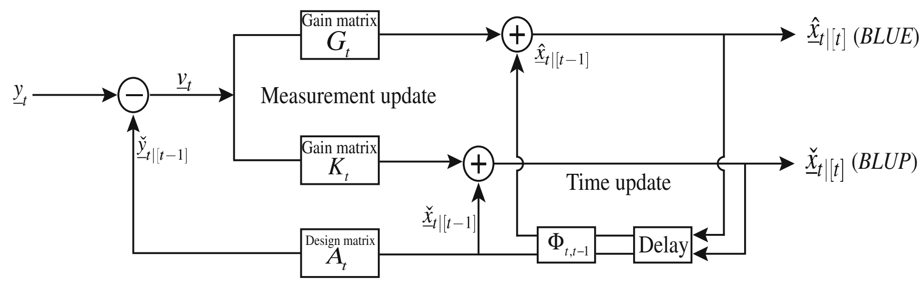

Fig. 1 Block diagram of the recursive BLUE–BLUP method with measurement and time updates

whereas the gain is computed from the error covariance matrix $C_{t|[t-1]}$.

As another important difference, note that in contrast to the BLUP recursion, the BLUE recursion cannot stand on its own. It requires the predicted residuals $\underline{v}_t$ and therefore the BLUP $\check{\underline{x}}_{t|[t-1]}$. Figure 1 shows the block diagram of the BLUE–BLUP recursion. The input is $\underline{y}_t$ and the outputs are the BLUE $\hat{\underline{x}}_{t|[t]}$ and the BLUP $\check{\underline{x}}_{t|[t]}$. The block diagram shows that estimation and prediction have the time update in common, but not the measurement update. The two measurement updates are fed with the same predicted residual, but have different gains. Their gain matrices are related as

$$G_t = C_{t|[t-1]} P^{-1}_{t|[t-1]} K_t \tag{63}$$

Let us now compare the two recursions with regards to their need of knowing $Q_{x_0x_0}$, the variance matrix of the initial state-vector $\underline{x}_0$. We already pointed out that this matrix does not play a role at all in the BLUP-recursion (cf. Theorem 2 (a)). It does however seem to play a role in the BLUE recursion, as it shows up in its initialization (60). A closer study of the mechanism of the BLUE recursion shows however that $Q_{x_0x_0}$ has also no effect on the outcomes of the BLUE. The gain matrix of the BLUE recursion is namely not driven by the variance matrix $Q_{t|[t-1]}$, but by the error covariance matrix $C_{t|[t-1]}$, which itself does not depend on $Q_{x_0x_0}$ (cf. Lemma 9). Hence, the only role played by $Q_{x_0x_0}$ lies in describing how the uncertainty of $\underline{x}_0$ contributes to the uncertainty of the estimators at the various epochs. Working with a model with unknown, but deterministic initial state-vector, i.e. $Q_{x_0x_0} = 0$, will therefore produce the same state-vector estimate as obtained when working with a random initial state-vector with unknown mean. Only the variance matrices of the two solutions will differ, since the latter is impacted by the randomness of the initial state-vector.

A further comparison between the two recursions shows that the difference between the BLUE and the BLUP is only driven by the system noise. Since both have the same initialization and the same time update, the difference between the two only starts to be felt in the measurement update of epoch $t = 1$. The measurement updates differ, since the gain matrices differ, $G_t \neq K_t$ (cf. 63). These gain matrices are the same however, in case $C_{t|[t-1]} = P_{t|[t-1]}$, which is the case when the system noise is absent. We therefore have the following result.

Corollary 5 (BLUE = BLUP) *The recursive BLUP becomes identical to the recursive BLUE, in case system noise is absent, i.e. if $S_t = 0$ for all t, then*

$$\hat{\underline{x}}_{t|[t]} = \check{\underline{x}}_{t|[t]} \quad \text{and} \quad \hat{\underline{x}}_{t|[t-1]} = \check{\underline{x}}_{t|[t-1]} \tag{64}$$

for all t.

Thus, in all cases where system noise is present the recursive BLUE will give an output different from that of the recursive BLUP.

5.3 Recursive BLUE–BLUP

Since the BLUE and BLUP recursions have the same structure and are both based on the same predicted residuals, one can combine them into one recursion. For that purpose, we denote the BLUE–BLUP state-vector and its error variance matrix as

$$\tilde{\underline{x}}_{t|[t-1]} = \begin{bmatrix} \hat{\underline{x}}_{t|[t-1]} \\ \check{\underline{x}}_{t|[t-1]} \end{bmatrix}, \quad \tilde{P}_{t|[t-1]} = \begin{bmatrix} Q_{t|[t-1]} & C_{t|[t-1]} \\ C^T_{t|[t-1]} & P_{t|[t-1]} \end{bmatrix} \tag{65}$$

with a likewise definition for $\tilde{\underline{x}}_{t|[t]}$ and $\tilde{P}_{t|[t]}$. Thus combining the recursions, the combined results of Theorems 2 (a) and (b) can be summarized as follows.

Theorem 2 (Recursive BLUE–BLUP) *The three steps of the BLUE–BLUP recursion are given as*

$$\textit{Initialization:} \quad \begin{aligned} \tilde{\underline{x}}_{0|0} &= E\hat{\underline{x}}_{0|0}, \\ \tilde{P}_{0|0} &= E P_{0|0} E^T + \tilde{Q}_{x_0x_0} \end{aligned} \tag{66}$$

with $E = [I_n, I_n]^T$ and $\tilde{Q}_{x_0x_0} = \text{blockdiag}(Q_{x_0x_0}, 0)$.

$$\textit{Time update:} \quad \begin{aligned} \tilde{\underline{x}}_{t|[t-1]} &= \tilde{\Phi}_{t,t-1}\tilde{\underline{x}}_{t-1|[t-1]}, \\ \tilde{P}_{t|[t-1]} &= \tilde{\Phi}_{t,t-1}\tilde{P}_{t-1|[t-1]}\tilde{\Phi}^T_{t,t-1} + \tilde{S}_t \end{aligned} \tag{67}$$

with transition matrix $\tilde{\Phi}_{t,t-1} = \text{blockdiag}(\Phi_{t,t-1}, \Phi_{t,t-1})$ and system noise variance matrix $\tilde{S}_t = \text{blockdiag}(0, S_t)$.

Measurement update:

$$\begin{aligned} \tilde{\underline{x}}_{t|[t]} &= \tilde{\underline{x}}_{t|[t-1]} + \tilde{K}_t \underline{v}_t, \\ \tilde{P}_{t|[t]} &= \tilde{P}_{t|[t-1]} - \tilde{K}_t Q_{v_tv_t} \tilde{K}^T_t \end{aligned} \tag{68}$$

Table 3 The three steps of the BLUE–BLUP recursion compared

	Estimation	Prediction
Initialization	$\hat{\underline{x}}_{0\|0} = (A_0^T R_0^{-1} A_0)^{-1} A_0^T R_0^{-1} \underline{y}_0$	$\check{\underline{x}}_{0\|0} = \hat{\underline{x}}_{0\|0} = (A_0^T R_0^{-1} A_0)^{-1} A_0^T R_0^{-1} \underline{y}_0$
	$Q_{0\|0} = Q_{x_0x_0} + (A_0^T R_0^{-1} A_0)^{-1}$	$P_{0\|0} = (A_0^T R_0^{-1} A_0)^{-1}$
Time update	$\hat{\underline{x}}_{t\|[t-1]} = \Phi_{t,t-1}\hat{\underline{x}}_{t-1\|[t-1]}$	$\check{\underline{x}}_{t\|[t-1]} = \Phi_{t,t-1}\check{\underline{x}}_{t-1\|[t-1]}$
	$Q_{t\|[t-1]} = \Phi_{t,t-1} Q_{t-1\|[t-1]} \Phi_{t,t-1}^T$	$P_{t\|[t-1]} = \Phi_{t,t-1} P_{t-1\|[t-1]} \Phi_{t,t-1}^T + S_t$
Measurement update	$\hat{\underline{x}}_{t\|[t]} = \hat{\underline{x}}_{t\|[t-1]} + C_{t\|[t-1]} A_t^T Q_{v_tv_t}^{-1} \underline{v}_t$	$\check{\underline{x}}_{t\|[t]} = \check{\underline{x}}_{t\|[t-1]} + P_{t\|[t-1]} A_t^T Q_{v_tv_t}^{-1} \underline{v}_t$
	$Q_{t\|[t]} = Q_{t\|[t-1]} - C_{t\|[t-1]} A_t^T Q_{v_tv_t}^{-1} A_t C_{t\|[t-1]}^T$	$P_{t\|[t]} = P_{t\|[t-1]} - P_{t\|[t-1]} A_t^T Q_{v_tv_t}^{-1} A_t P_{t\|[t-1]}$

with predicted residual $\underline{v}_t = \underline{y}_t - \tilde{A}_t \tilde{\underline{x}}_{t|[t-1]}$, $\tilde{A}_t = A_t[0, I_n]$, *and gain matrix* $\tilde{K}_t = \tilde{P}_{t|[t-1]} \tilde{A}_t^T Q_{v_tv_t}^{-1}$.

This result shows how the recursive BLUE and the recursive BLUP can be mechanized into one single recursion. This result is therefore the recursive formulation of the BLUE–BLUP expression given in Theorem 1. With this extention of the standard 'Kalman filter' theory, we are thus also able to recursively compute the best estimate of the unknown mean state-vector, instead of only the best prediction of the random state-vector outcome.

In analogy with Kalman filter-based smoothing, it also possible to develop the BLUE–BLUP smoothing solution. For the BLUE part, smoothing is rather straightforward, since $\hat{\underline{x}}_{t|[s]} = \Phi_{t,s}\hat{\underline{x}}_{s|[s]}$ and $Q_{t|[s]} = \Phi_{t,s} Q_{s|[s]} \Phi_{t,s}^T$. For the BLUP part, the smoothing will resemble the standard smoothing methods, like fixed-point, fixed-interval or fixed-lag smoothing, see e.g. (Gelb 1974; Maybeck 1979; Jazwinski 1991; Gibbs 2011).

6 Summary and conclusions

In this contribution, the BLUE–BLUP recursion of the partitioned measurement and dynamic models was introduced (see Table 3). It extends 'Kalman-filter' theory by replacing the BLP approach with the BLUP, thereby relaxing the assumptions on the state-vector means. It was argued that the BLUP approach is often more appropriate, since in many, if not most, applications the means of the state-vectors are indeed unknown.

The recursive BLUP was derived from first principles, thereby making use of an earlier derived decomposition of the BLUP into misclosures and any LUP. The role of the misclosures was emphasized and it was shown how they form the basis for constructing the predicted residuals. It was also shown how the recursive BLUP, as a consequence of the relaxing assumption on the state-vector means, does away with the need of having to specify the initial state-vector variance matrix.

Next to the recursive BLUP, we introduced, for the same model, the recursive BLUE. This extension is new and another consequence of assuming the state-vector means unknown. In the standard Kalman filter set-up with known state-vector means, such difference between estimation and prediction does not occur since one is then only left with BLP instead of with BLUP of the state-vectors.

Finally, it was shown how the two intertwined recursions can be combined into one general BLUE–BLUP recursion. The recursion outputs for every epoch, in parallel, the BLUP for the random state-vector and the BLUE for the mean of the state-vector (cf. the block diagram of Fig. 1).

Acknowledgments P.J.G. Teunissen is the recipient of an Australian Research Council Federation Fellowship (project number FF0883188). The research of A. Khodabandeh was carried out whilst a Curtin International Research Scholar at Curtin's GNSS Research Centre. All this support is gratefully acknowledged.

References

Anderson BDO, Moore JB (1979) Optimal filtering, vol 11. Prentice-hall, Englewood Cliffs
Ansley CF, Kohn R (1985) Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions. Ann Stat 13(4):1286–1316
Bar-Shalom Y, Li X (1993) Estimation and tracking—principles, techniques, and software. Artech House, Inc., Norwood
Bode H, Shannon C (1950) A simplified derivation of linear least square smoothing and prediction theory. Proc IRE 38(4):417–425
Brammer K, Siffling G (1989) Kalman-Bucy filters. Artech House, Norwood
Candy J (1986) Signal processing: model based approach. McGraw-Hill, Inc, New York
Gelb A (1974) Applied optimal estimation. MIT Press, Cambridge
Gibbs B (2011) Advanced Kalman filtering, least-squares and modeling: a practical handbook. Wiley, New York
Goldberger A (1962) Best linear unbiased prediction in the generalized linear regression model. J Am Stat Assoc 57(298):369–375
Grewal MS, Andrews AP (2008) Kalman filtering: theory and practice using MATLAB, 3rd edn. Wiley, New York
Harvey AC, Phillips GDA (1979) Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. Biometrika 66(1):49–58
Jazwinski A (1991) Stochastic processes and filtering theory. Dover Publications, New York
de Jong P (1991) The diffuse Kalman filter. Ann Stat 19(2):1073–1083
Kailath T (1968) An innovations approach to least-squares estimation—part I: Linear filtering in additive white noise. IEEE Trans Autom Control 13(6):646–655

Kailath T (1981) Lectures on Wiener and Kalman filtering. Springer, Berlin

Kailath T, Sayed AH, Hassibi B (2000) Linear estimation. Prentice-Hall, Englewood Cliffs

Kalman RE (1960) A new approach to linear filtering and prediction problems. J Basic Eng 82(1):35–45

Maybeck P (1979) Stochastic models, estimation, and control, vol 1. Academic Press, Waltham, republished 1994

Sanso F (1986) Statistical methods in physical geodesy. In: Sunkel H (ed) Mathematical and numerical techniques in physical geodesy. Lecture notes in earth sciences, vol 7. Springer, Berlin, pp 49–155

Simon D (2006) Optimal state estimation: Kalman, H [infinity] and nonlinear approaches. Wiley, New York

Sorenson HW (1966) Kalman filtering techniques. In: Leondes CT (ed) Advances in control systems: theory and applications, vol 3. pp 219–292

Stark H, Woods J (1986) Probability, random processes, and estimation theory for engineers. Prentice-Hall, Englewood Cliffs

Teunissen PJG (2000) Adjustment theory: an introduction. In: Series on Mathematical Geodesy and Positioning. Delft University Press

Teunissen PJG (2007) Best prediction in linear models with mixed integer/real unknowns: theory and application. J Geod 81(12):759–780

Teunissen PJG (2008) On a stronger-than-best property for best prediction. J Geod 82(3):167–175

Teunissen PJG, Simons DG, Tiberius CCJM (2005) Probability and observation theory. Delft University, Faculty of Aerospace Engineering, Delft University of Technology, lecture notes AE2-E01

Tienstra J (1956) Theory of the adjustment of normally distributed observation. Argus, Amsterdam

Zadeh LA, Ragazzini JR (1950) An extension of Wiener's theory of prediction. J Appl Phys 21(7):645–655