Kalman Filter Algorithm for the joint Processing of GNSS PPP and Accelerometer Data, EEW parameters from the *unbiased* displacement time-series

Andreas Rosenberger [*], Simon Banville [†], Paul Collins [‡] Joe Henton [§], Eli Ferguson [¶]

V0.9, Documentation for ONC-EEW, Dec 2016 - Apr. 2018

Changes to this document

Changed, V0.7: Merged the "Kalman" and "EEW parameter..." documents for easier cross-reference. EEW parameters are now discussed in section 6. Section 5.2.3 discusses potential bias in the PPP data. AR 17-09-08

Added, V0.8: Section 5.1.3 addressing the potential problem of acceleration data gaps. AR, EF, PC 18-04-16

Corrected V0.9: Section 6.3, the suggested time interval of $60s$ was contradicting recommendations make in section 5.2.3 to actually use $600s$ for the computation of mean and standard deviation, AR 10-07-17

1 Overview

The filter owes its name to R. E. Kalman at the Institute for Advanced Studies in Baltimore (Kalman, 1960), developed well before the start of the digital age. It has since found numerous applications in control, signal estimation and general filtering problems. Although similar proposals have been made before (c.f. Mayhew, 1999) in a different context, the proposal to apply a Kalman filter to the problem of combining observations from the Global Navigation Satellite Systems (GNSS) with data from a three axis accelerometer was initially proposed by Smyth and Wu, 2007. The general idea of the Kalman filter is to predict a system's behaviour from incomplete observations. The problem here is to predict high resolution displacement and velocity from acceleration and Precise Point Position (PPP) data which are available at different times and different rates from the accelerometer and corrected GNSS data respectively. The idea to apply this to real-time seismology and earthquake early warning (EEW) goes back to Bock, Melgar,

[*]andreas@arescon.com
[†]simon.banville@canada.ca
[‡]paul.collins@canada.ca
[§]joe.henton@canada.ca
[¶]elif@uvic.ca

and Crowell, 2011 and has been validated in several studies (Melgar et al., 2013; Li, 2015; Niu and Xu, 2014) where various data-sets from large earthquakes were processed off-line, after the event.

2 The Basic Kalman Equations

Given acceleration time-series, the task is to generate velocity and displacement estimates, which is in general a two-step integration process. However, each integration would require that bias and the very low-frequency components of the acceleration and velocity signal are removed prior to each integration step. Bias would let the integrated signal grow out of bounds over time and small errors in the very low-frequency components would be disproportionally amplified. The goal is to force the bias and low-frequency components in the combined displacement time-series to largely follow the PPP data while retaining the high-frequency components derived from double integration of the acceleration time-series. The resulting *unbiased displacement time-series* should thus have the characteristics from the best of each of the two worlds.

2.1 The Estimation Problem

In the Kalman approach the problem is cast as a recursive estimation problem, values of displacement and velocity from one step back in time (d_{t-1}, v_{t-1}) are used to estimate the state of the system in terms of displacement and velocity at the current time (d_t, v_t), given a measured acceleration value (a_m), a sample from a single channel of the accelerometer, and an estimation error term (η_a).

This is cast as linear problem, in explicit matrix notation:

$$\begin{bmatrix} d_t \\ v_t \end{bmatrix} = \begin{bmatrix} 1 & \delta t_a \\ 0 & 1 \end{bmatrix} \begin{bmatrix} d_{t-1} \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} {\delta t_a}^2/2 \\ \delta t_a \end{bmatrix} \cdot a_m + \begin{bmatrix} {\delta t_a}^2/2 \\ \delta t_a \end{bmatrix} \cdot \eta_a \tag{2.1}$$

where δt_a is the sampling time interval of the acceleration time series, and its appearance in the matrix and vector terms performs the respective single and double integrations over one time interval.

A second equation describes the measured displacement d_m from the GNSS PPP derived data

$$d_m = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} d_{t-1} \\ v_{t-1} \end{bmatrix} + \epsilon_d \tag{2.2}$$

in terms of the estimates d_{t-1} and v_{t-1} and an estimation error ϵ_d.

The key problem then is to synchronize and combine those two equations in a way which will minimize the estimation errors in the velocity and displacement time series over time.

In compact form, as used in Smyth and Wu, 2007, equations 2.1 and 2.2 are written as:

$$\hat{\mathbf{x}}_\mathbf{t} = \mathbf{A}\mathbf{x}_{\mathbf{t-1}} + \mathbf{B}a_m + \mathbf{w} \tag{2.3}$$

$$d_m = \mathbf{H}\mathbf{x}_{\mathbf{t-1}} + \epsilon_\mathbf{d} \tag{2.4}$$

In equation 2.3 $\hat{\mathbf{x}}_\mathbf{t}$ is seen as an *a-priori* estimate and equation 2.4 is used to formulate an *a-posteriori* estimation error $d_m - \mathbf{H}\mathbf{x}_{\mathbf{t-1}}$ which is used to improve the initial estimate $\hat{\mathbf{x}}_\mathbf{t}$

$$\mathbf{x_t} = \hat{\mathbf{x}}_\mathbf{t} + \mathbf{K_t}(d_m - \mathbf{H}\mathbf{x_{t-1}}). \tag{2.5}$$

Here $\mathbf{K_t}$ is the so-called Kalman gain, $\mathbf{K_t} = \begin{bmatrix} K_t^d \\ K_t^v \end{bmatrix}$ which needs additional quantities to be computed. The first quantity is the covariance matrix estimate using the accelerometer noise variance, according to Smyth and Wu, 2007:

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{3}q\delta {t_a}^3 & \frac{1}{2}q\delta {t_a}^2 \\ \\ \frac{1}{2}q\delta {t_a}^2 & q\delta t_a \end{bmatrix} \tag{2.6}$$

where q stands for the variance of the accelerometer time series. This version of the matrix Q is according to Bar-Shalom, Li, and Kirubarajan, 2001 (their equation 6.3.2-4, p.273) not correct. It should be

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{4}q\delta {t_a}^4 & \frac{1}{2}q\delta {t_a}^3 \\ \\ \frac{1}{2}q\delta {t_a}^3 & q\delta {t_a}^2 \end{bmatrix} \tag{2.7}$$

However, the difference is most likely without consequences for the algorithm's performance and is briefly discussed in section 3.1.

The second quantity is a covariance estimate of the errors in equations 2.1 (or 2.3) which comes in an *a-priori* and an *a-posteriori* version.
The *a-priori* version is:

$$\hat{\mathbf{P}}_\mathbf{t} = \mathbf{A}\mathbf{P_{t-1}}\mathbf{A^T} + \mathbf{Q} \tag{2.8}$$

and the *a-posteriori* version becomes:

$$\mathbf{P_t} = (\mathbf{I} - \mathbf{K_t}\mathbf{H})\hat{\mathbf{P}}_\mathbf{t}, \tag{2.9}$$

where $\mathbf{I}$ is the 2×2 identity matrix.

The Kalman gain vector then becomes:

$$\mathbf{K_t} = \hat{\mathbf{P}}_\mathbf{t}\mathbf{H^T}\left[\mathbf{H}\hat{\mathbf{P}}_\mathbf{t}\mathbf{H^T} + \mathbf{r_d}\right]^{-1}, \tag{2.10}$$

where $\mathbf{r_d} = \frac{\sigma_t}{\delta t_d}$, σ_t is the variance and δt_d is the sampling time interval of the PPP data.

3 Multi-rate Algorithm

Obviously equations 2.3 to 2.10 have to be executed in proper order and acceleration data and GNSS derived PPP data are available with different sampling rates. Therefore equation 2.5, giving the *a-posteriori* estimate of displacement and velocity can only be executed when a new PPP sample is available. The same applies to equation 2.9 where the *a-posteriori* covariance matrix is updated.

The final algorithm can be stated as follows:

For every acceleration sample a_m execute:

$$\mathbf{x_t} = \mathbf{A}\mathbf{x_{t-1}} + \mathbf{B}a_m \tag{3.1}$$

$$\mathbf{P_t} = \mathbf{A}\mathbf{P_{t-1}}\mathbf{A^T} + \mathbf{Q} \tag{3.2}$$

If a PPP sample d_m is available execute instead:

$$\hat{\mathbf{x}}_\mathbf{t} = \mathbf{A}\mathbf{x_{t-1}} + \mathbf{B}a_m \tag{3.3}$$

$$\hat{\mathbf{P}}_\mathbf{t} = \mathbf{A}\mathbf{P_{t-1}}\mathbf{A^T} + \mathbf{Q} \tag{3.4}$$

$$\mathbf{K_t} = \hat{\mathbf{P}}_\mathbf{t}\mathbf{H^T}\left[\mathbf{H}\hat{\mathbf{P}}_\mathbf{t}\mathbf{H^T} + \mathbf{r_d}\right]^{-1} \tag{3.5}$$

$$\mathbf{x_t} = \hat{\mathbf{x}}_\mathbf{t} + \mathbf{K_t}(d_m - \mathbf{H}\mathbf{x_{t-1}}) \tag{3.6}$$

$$\mathbf{P_t} = (\mathbf{I} - \mathbf{K_t}\mathbf{H})\hat{\mathbf{P}}_\mathbf{t} \tag{3.7}$$

The vector $\mathbf{x_t} = \left[\begin{smallmatrix} d_t \\ v_t \end{smallmatrix}\right]$ then holds the current samples of the unbiased displacement and velocity time-series.

Initialization

The constant matrices $\mathbf{A}$ and $\mathbf{B}$ are as in equation 2.1, $\mathbf{H}$ is the constant row-vector from equation 2.2. The algorithm is initialized with $\mathbf{Q}$ as in equation 2.7 and

$$\mathbf{P_{t-1}} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$$

$$\mathbf{x_{t-1}} = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$$

.

$$\mathbf{r_d} = \frac{\sigma_\mathbf{t}}{\delta\mathbf{t_d}},$$

where δt_d is the sampling time interval of the PPP time-series and σ_t is the estimated variance of the PPP data.

3.1 Variances of accelerometer and PPP data

So far the variances in the accelerometer and the PPP data have been treated as constant values which can be estimated as measurements of the noise from the respective time-series in the absence of a signal. This not a very realistic assumption since for example the accelerometer may get rotated or tilted in a strong earthquake which would change the variance as a measure of the accuracy of the data. Additionally, as indicated before, the correct version of the matrix Q according to Bar-Shalom, Li, and Kirubarajan, 2001

should be equation 2.7. As long as the variance of the accelerometer data q is seen as a *free* parameter, all that matters here is that Q is symmetric and positive.

The variance σ_t of the PPP data will vary over time, depending on the satellite constellation for example.

In the Kalman algorithm the assigned variances are measures of the (un-)reliability of the respective data. Saunders et al., 2016, for example treat the accelerometer variance values q in 2.6 or 2.7 simply as *Kalman filter weights* and assign "variance multipliers", factors of 10, 100, 1000 to the accelerometer variance in order to "optimize" the impact of noisy accelerometer values on the final displacement time series. In essence the accelerometer variance (and in consequence Q) becomes a *fudge factor* which may have to be determined on a site-by-site basis. Additionally the variance of the GNSS PPP data can be estimated with the available corrections data stream used by the PPP engine and the variance σ_t is reported with each sample of the respective PPP channel. Thus $\mathbf{r_d}$ in equation 2.10 will have to be updated on a sample by sample basis for each reported σ_t.

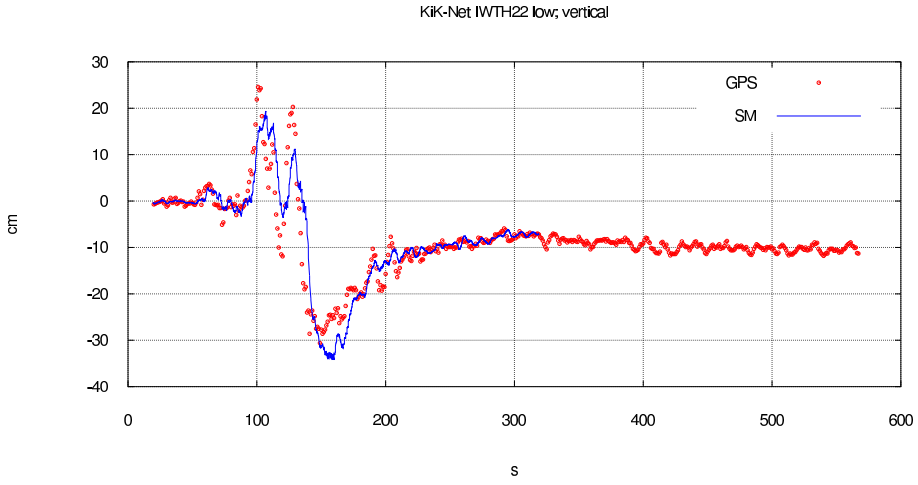

Figure 1: Unbiased vertical displacement (blue line) from the IWTH22 Kik-Net accelerometer and GNSS station MITSU-070 (red dots). Data are from the 2011-03-09 Tohoku earthquake.

4 Implementation

Little is known about the long term numerical stability of the computations for this particular Kalman algorithm in a true on-line, real-time application. However, most calculations are relatively simple and the inverse term in equation 2.10 actually evaluates to a scalar.

Implicit in the formulation of the joint on-line processing of acceleration and PPP data is, that both data streams are perfectly synchronized, they cannot be off-set in time against one another. Assuming that this cannot be guaranteed a-priori, a possible time-offset PPP versus acceleration data would have to be determined experimentally and compensated for, delaying the earlier data stream before submitting the data to the Kalman algorithm.

The accelerometer data will invariably exhibit a small bias and it is suggested to apply a digital high-pass filter with a cut-off frequency of $0.075Hz$ ($T \approx 13s$) to all three channels before they are passed to

the Kalman algorithm. The combination with the PPP data will reconstruct the low frequencies (and bias from co-seismic displacement). This should, however, be verified in an experiment.

The *un-biased displacement and velocity time-series* will have the same sampling rate as the acceleration data. It is suggested (see Crowell et al., 2013) to calculate the P-wave peak displacement amplitude P_d, as described in Rosenberger, 2014, directly from the *un-biased displacements* since their low-frequency spectrum should be intact and thus help to avoid the saturation effect in magnitude estimates for large earthquakes. It is also suggested that the resulting un-biased velocity and displacement data from the P-wave can be used in a more stable way (Lior, Ziv, and Madariaga, 2016) to calculate the predominant period parameter τ_p (*I would have to look into that*).

From the P-wave detection time on, peak ground acceleration (PGA), peak ground velocity (PGV) and peak ground displacement (PGD) should be reported down-stream once a second (possibly even faster). Future down-stream applications (re-construction of the earthquake source mechanism, for example) may require additional parameters to be generated.

4.1 Configurable Parameters

- Accelerometer variance fudge factor (eq. 2.7, parameter q)

  q is regarded as a constant value within the Kalman context, however, it does not actually relate to the physical/statistical noise in the accelerometer data (Saunders et al., 2016). We may need to adjust this value on a site by site basis. It is suggested to introduce a multiplier, a single factor α for each of the accelerometer channels so that $q = \alpha\sigma^2$, where σ is the assumed accelerometer noise standard deviation.

- Time interval for monitoring PPP standard deviation (section 6 eqs. 6.1, 6.2)

  The time interval from which standard deviation is computed, suggested is an initial value of 1 minute (or 60 samples), should also be configurable. One may have to increase this value to possibly several minutes.

4.1.1 Monitoring of long-term numerical stability

It has been suggested that monitoring the symmetry of matrix P_t (eqs. 2.8, 2.9) is a means to assess potential numerical stability issues. Thus checking $P_t(1,2) == P_t(2,1)$ (with a small numerical tolerance) could signal problems we would have to address.

4.1.2 Convergence

Convergence of the Kalman algorithm can be estimated by monitoring the value $P_t(1,1)$. Convergence is established when $P_t(1,1)$ has decreased to a small value that remains more or less constant.

Proposed Processing Scheme

An overview of the envisioned processing scheme is given in figure 2.

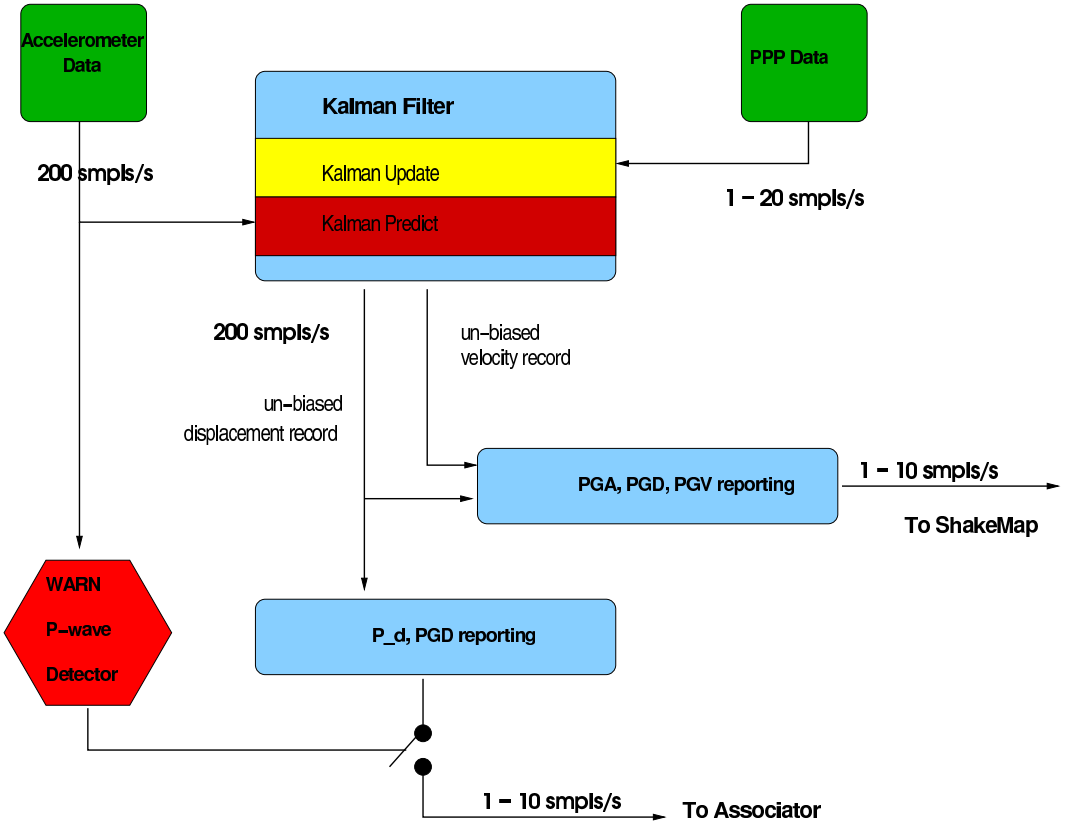

Figure 2: Kalman processing and the WARN P-wave detection

5 On-site Monitoring of Data Quality in Accelerometer and PPP Data

5.1 Accelerometer/Digitizer, on-site Computer System

The accelerometer installations are either TitanSMA, including a digitizer or TitanEA instruments connected to a Centaur digitizer. In both cases digital data are pre-processed by John Dorocizc' software which also applies a 0.075Hz high-pass filter which should remove bias from the acceleration time series. The same software also provides P- and S-wave detections which are signalled to the Kalman algorithms. By default the Titans or Centaurs synchronize to Universal Time Coordinated (UTC) with their own independent GPS receivers.

The on-site computer hosting all processing software is synchronizing to UTC by means of the network time protocol (NTP).

5.1.1 Possible error conditions

In normal operations the most likely error to occur is a missed P-wave detection, the instrument may just be too far from the epicenter, the Kalman algorithms would then be unable to report P_d from the unbiased displacement time series (eq. 6.3). The S-wave, due to its larger amplitude, may still be detected and should trigger the computation of P_{PGD} (eq. 6.5).

The discrimination of P- and S-waves however can fail, assigning the wrong wave type to a detection. This type of error would be detected in the associator/correlator and the reported displacement values would be dismissed.

5.1.2 Time synchronization

The Titan accelerometers could loose time synchronization, due to a failure of the Titan's or Centaur's independent GPS. Apparently a pending firm-ware upgrade for the Titan-SMA will use network time protocol (NTP) servers as a back-up.

It still needs to be confirmed that a Nanometrics Centaur digitizer can do the same and that reliable NTP-servers are pre-configured and are reachable via the Internet connections.

The PPP-Engine transmits three data streams from different sets of corrections to the Kalman algorithms, based on broadcast orbits (BO), floating point ambiguity resolution (FpAR) and integer ambiguity resolution (IntAR). The PPP-Engine is synchronized to GPS time and thus offset from UTC by a number of leap-seconds (currently GPS time leads UTC by 18 seconds). This number will change over time. Leap seconds are added to UTC about every 18 months, normally added either on December 31 or on June 30 but any other last day of the month would also be possible. UTC becomes discontinuous whenever a leap-second is added which poses a problem since a UTC synchronized accelerometer would nominally produce excess data with invalid time-stamps whenever the UTC clock is halted for one second.

A future version of the PPP-engine will provide data with UTC time-stamps also to accommodate timing standards used in seismology. In consequence, whenever a leap-second is added to UTC in the

future, there will be two one second data sets with identical time stamps. This would also affect potential heart-beat reporting and downstream data and quality management needs to be prepared to handle those.

5.1.3 Missing Segments of Acceleration Data

A hopefully rare condition could be a missing segment of acceleration data, be it due to a brief breakdown in transmission from the digitizer or a power failure which only affects the accelerometer. An experiment with the data from the 2011-03-09 Tohoku earthquake (as in figure 1) suggests that updating the Kalman equations 3.1 to 3.7 with a_m (eqs. 3.1 and 3.3) set to zero is a valid strategy as long as PPP samples are still are being supplied regularly. This will keep matrix $\mathbf{P_t}$ in a state of convergence and once acceleration data become available again the unbiased displacement time-series recovers relatively fast. Figures 1, 3, 4 and 5 were generated driving the Kalman algorithm with a PPP variance of $1cm^2$ and an acceleration variance of $0.01cm^2/s^4$. Surprisingly, increasing the acceleration variance ten-fold during the data-gap has little influence on the general outcome.

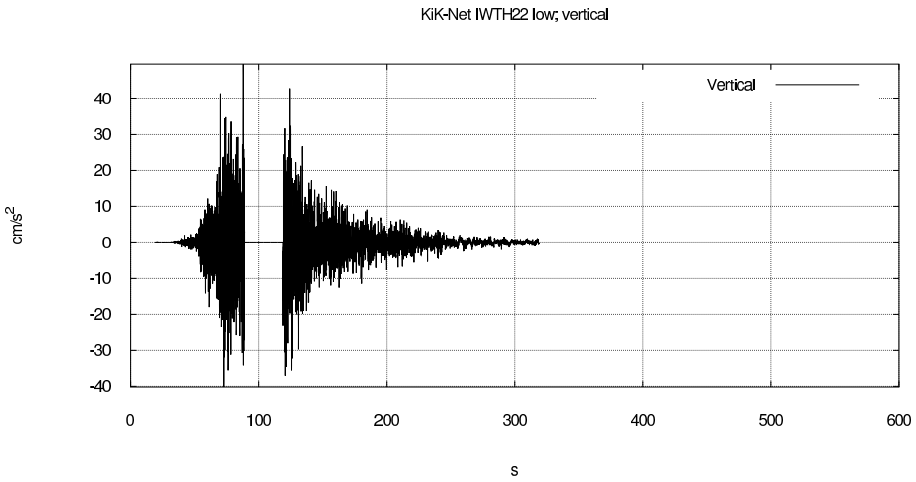

Figure 3: IWTH22 Kik-Net vertical accelerometer with data-gap from 90 to 120 seconds

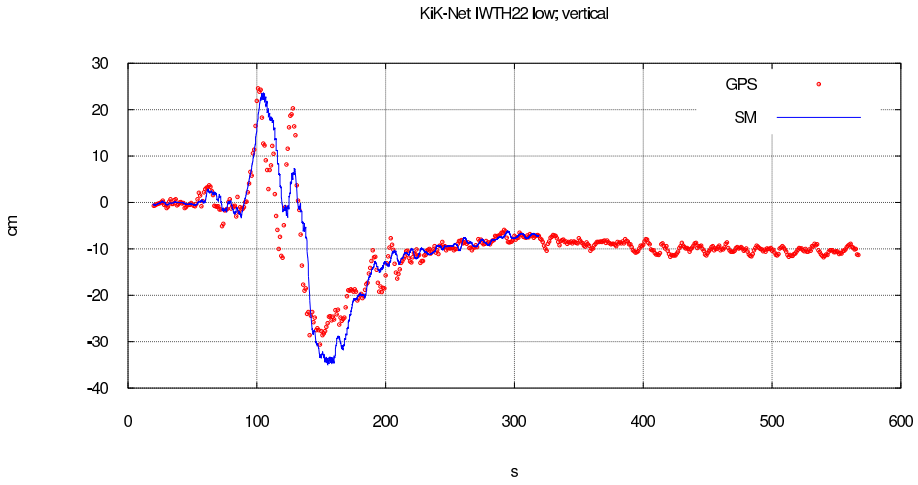

Figure 4: GPS PPP data and unbiased displacements, based on acceleration with data gap, the result with continuous data is shown in figure 1.

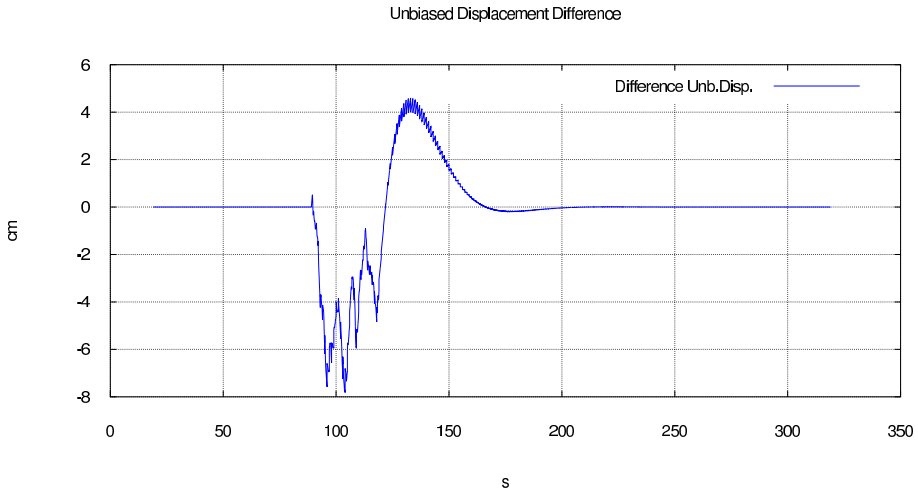

Figure 5: Difference of unbiased displacement with and without acceleration data gap

5.2 PPP-data streams

5.2.1 Latencies in the PPP-data

Normally PPP-data from all three sets of corrections will have latencies in the order of a few seconds. Latencies should be monitored and if they grow to values over ten seconds, this condition should be reported to ONC's data management system.

If latencies exceed the maximum life time of buffered accelerometer data, suggested is a buffer holding 15 seconds of accelerometer data, the respective Kalman algorithm needs to be suspended and re-set to initial conditions (see section 3). When the PPP data resume, re-convergence can be monitored as described in section 4.1.2. When regular latencies of several seconds occur in the PPP data, data delivery may resume initially at a much higher rate than one sample per second.

5.2.2 Limited availability of PPP data

All three PPP streams may degrade with unfavourable satellite constellations and FpAR-PPP and IntAR-PPP may become un-available if inbound communications break down. The Kalman algorithm would then have to fall back on BO-PPP, the corresponding unbiased displacement time-series will have larger Sigmas (eq. 2.10).

If FpAR-PPP and/or IntAR-PPP data streams resume after an interruption, the associated Kalman filters would have to be re-set to initial conditions (see section 3). Re-convergence can be monitored as described in section 4.1.2.

The respective unbiased displacement data become valid only after re-convergence , no P_d or P_{PGD} values should be reported before (assuming that an earthquake is detected during convergence).

5.2.3 Bias in the PPP data

The output PPP data will be periodically re-calibrated to the GNSS receiver *nominal* position. However, there will be a small bias in the relative position data which will vary slowly over time (with tidal frequencies and below). Any significant bias would transfer through the Kalman filter and potentially up-set the parameter computations needed for earthquake magnitude estimates (see sections 6.4 and 6.5).

It is therefore suggested to remove any bias from the PPP data *before* they are submitted to the Kalman filter.

This can be accomplished by computing a moving average as the mean value of the data over a nominal time interval. The moving average of a signal $d(t)$ at time t is approximated as

$$\begin{aligned} \bar{d}(t) &= \frac{1}{N}\left(\sum_{i=t-N+1}^{t-1} d(i) \quad + d(t)\right) && (5.1)\\ &\approx \frac{N-1}{N}\bar{d}(t-1) + \frac{1}{N}d(t) && (5.2)\\ &= \alpha\bar{d}(t-1) + (1-\alpha)d(t), && (5.3) \end{aligned}$$

with $\alpha = \frac{N-1}{N}$, where $N = \Delta T/\delta t$ is the window size in samples, ΔT the corresponding window lenght in seconds and δt is the sampling time interval. The moving average is updated whenever a new

PPP sample $d(t)$ becomes available.

To capture the slowly varying bias in the 1Hz PPP data of each channel a time-window of 600 seconds would be a starting point. The time window should be a configurable value.

The value $\bar{d}(t)$ then represents the current bias of each channel and can be subtracted from the data before they enter the Kalman computations.

Updating $\bar{d}(t)$ should be suspended as soon as the accelerometer detects an event. We are interested in the conditions immediately before the event.

5.2.4 Quality Indicators in the PPP Data Structures

1. Field 28 : nsat_use, normally nsat_use > 4, nsat_use $= 0$ indicates **no solution**

2. Fields 13, 14, 15: sN, sE, sh, normally sN and sE $< 0.02m$ and sh $< 0.03m$ would indicate that the solution has formally converged and should be precise enough to use. However those values will vary with the current satellite constellation and *may also be site specific*.

3. Integer Ambiguity Resolution only, Field 34: ffix_amb, a change from ffix_amb < 4 to ffix_amb > 4 may introduce a spike in the position time-series and a change in the mean positions. If the Kalman filter(s) work as intended, a transient in the PPP time-series should have only a small effect on the unbiased displacement time series.

6 EEW Parameters from the unbiased displacement time-series

6.1 Overview

There are still only a few seismo-geodetic networks in operation. In consequence there are only few suitable accelerometer and GNSS data-sets from medium to large earthquakes available today and this in turn limits the number of studies which have been carried out to investigate their theoretical performance in the context of Earthquake Early Warning (EEW). Even in Japan the accelerometers of KiK-net and K-NET (Aoi, Kunugi, and Fujiwara, 2004) are not directly collocated with the GNSS antennas of GEONET (www.gsi.go.jp, Sagiya, 2004). Due to the high station density of KiK-net and K-NET an accelerometer can be found though in a few kilometres distance from a GEONET site.

Most studies investigating the joint processing of GNSS and accelerometer data have appeared after the $M9.0$ Tohoku-Oki event in 2011. Earlier studies (Crowell, Bock, and Squibb, 2009) investigated just the use of high-rate GNSS data alone for EEW. Emore et al., 2007 is often cited as one of the earliest publications proposing the joint processing of GNSS and strong-motion accelerometer data in the more general context of seismology. A *Kalman* filter (Section 1) is used to fuse high-rate GNSS *Precise Point Positioning* (PPP) data and acceleration time-series into what is called the *unbiased displacement* time-series.

Crowell et al., 2013 developed magnitude scaling relationships for observed P-wave peak displacements from five large earthquakes with magnitudes between $M5.4$ and $M9.0$ where GNSS-PPP and accelerometer data from "collocated" sites (accelerometer and GNSS within $4km$ distance) were available. They also developed scaling relationships for peak displacements over the whole duration of the earthquake which can be used to update magnitude estimates periodically until earthquake motions finally subside and a final static displacement is reached. A follow-up study (Crowell et al., 2016) used real and synthetic data to investigate scaling relationships for the 28 February 2001, $M_w6.8$ Nisqually earthquake.

6.2 Unbiased Displacement Time Series

Currently the PPP engine will supply three different time-series of $1Hz$ PPP data (in increasing order of accuracy):

1. based on broadcast orbits (BO PPP)
2. based on floating point ambiguity resolution (floatAR PPP)
3. based on integer ambiguity resolution (intAR PPP)

All three solutions will have different latencies and the only PPP time-series which is independent of correction streams coming in over the Internet is the broadcast orbits time-series since the associated corrections are transmitted with the GNSS signal. **Not all three PPP time-series will always be available**, integer-AR PPP being the most desirable due to its high accuracy.

Each of the three components, (north, east, vertical) of accelerometer and the respective PPP data has to be processed by one instance of the Kalman filter algorithm.

6.3 Quantifying Error in the Unbiased Displacements

The output of the Kalman filter will be affected from noise in the accelerometer data as well as in the PPP displacement data. The accelerometer noise contributions are difficult (if not impossible) to quantify, in particular during strong ground motions. It is therefore suggested to use the noise, usually quantified as standard deviation from the PPP displacement data, *in the absence of ground shaking*, as a proxy since they are not expected to change dramatically during an earthquake.

However, the standard deviations reported with the respective PPP displacements are *not* intended or suited to quantify the uncertainty of the actual PPP displacements. Their standard deviations will vary over time and have to be estimated directly from the discrete PPP displacement data $d(t)$.

Formally standard deviation is computed as

$$\sigma_t = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(d(i\cdot\delta t) - \bar{d}(t)\right)^2} \tag{6.1}$$

with

$$\bar{d}(t) = \frac{1}{N}\sum_{i=1}^{N} d(i\cdot\delta t) \tag{6.2}$$

being the mean value of the displacements $d(i\cdot\delta t)$, $i = 1, .., N$.

> *A stable mean value would already be available from the computation of $\bar{d}(t)$ as in equation 5.3.*

The question remains which value of N would be suitable to generate statistically robust and meaningful values for σ_t. It is *strongly recommended* to design the buffer size N as a configurable parameter. With current PPP sampling frequencies of 1 Hz a buffer of 600 seconds worth of data ($N = 600$) would be a starting point for experiments.

Standard deviations have to be updated periodically. An update time interval of 1 second is suggested. This would apply also with higher PPP sampling rates.

> Updating σ_t should be suspended as soon as the accelerometer detects an event. We are interested in the conditions immediately before the event.

6.4 P-wave Processing

Following Crowell et al., 2013, from the P-wave detection time on, one computes

$$P_d = max_{t=0}^{t=5s}\left[\sqrt{(d_N(t)^2 + d_E(t)^2)}\right] \tag{6.3}$$

over 5 seconds from the P-wave trigger time from the two *horizontal* components of *unbiased* displacement $d_N(t)$ and $d_E(t)$ from each of the available PPP solutions.

The last pre-event, σs (eq. 6.1), immediately before the P-wave detection time t_P, are reported as

$$\Sigma_d(t_P) = \sqrt{\left(\sigma^2_{t_{P_N}} + \sigma^2_{t_{P_E}}\right)} \tag{6.4}$$

The resulting three value pairs

- P_d^{BO}, $\Sigma_d^{BO}(t_P)$
- $P_d^{floatAR}$, $\Sigma_d^{floatAR}(t_P)$
- P_d^{intAR}, $\Sigma_d^{intAR}(t_P)$

are reported to the associator module with the time-stamp of the initial detection.

A possible extension would be to also report the maximum of the internal Kalman variances $\mathbf{P_t}(\mathbf{1},\mathbf{1})$ (eqn. 3.7, first row, first column element of $\mathbf{P_t}$). We should determine *experimentally* if those are useful in combination with the PPP standard deviations in order to quantify the reliability of the P_d values.

6.5 Updates over the full Duration of the Earthquake

Large earthquakes can have rupture durations of several hundred seconds. The dynamic displacement over the full duration was also used by Crowell et al., 2013 to develop a scaling relation based on observations of the peak ground displacement (PGD) computed from all three components as

$$P_{PGD}(T) = max_{t=0}^{T}\left[\sqrt{(d_N(t)^2 + d_E(t)^2 + d_Z(t)^2)}\right] \tag{6.5}$$

It is suggested to compute (update) $P_{PGD}(t)$ over 200 seconds from the P-wave or S-wave detection time $t = 0$ (a station in greater distance may have missed the P-wave) and report its current value $P_{PGD}(T)$ at 1 second time intervals together with the standard deviations of the PPP samples from equation 6.1 from immediately before the P-wave detection time t_P

$$\Sigma_{PGD}(t_P) = \sqrt{\left(\sigma^2_{t_{P_N}} + \sigma^2_{t_{P_E}} + \sigma^2_{t_{P_Z}}\right)} \tag{6.6}$$

As before reported are the three value pairs

- $P_{PGD}^{BO}(T)$, $\Sigma_{PGD}^{BO}(t_P)$
- $P_{PGD}^{floatAR}(T)$, $\Sigma_{PGD}^{floatAR}(t_P)$
- $P_{PGD}^{intAR}(T)$, $\Sigma_{PGD}^{intAR}(t_P)$

with the associated time-stamps T.

6.6 Heartbeat

The standard deviations associated with the PGD values from equation 6.6 could also serve as a heartbeat message. They are an indicator for the reliability of the different PPP solutions and would also show which of the three PPP solutions are available at the current instance in time.

7 Possible Extensions

Seismo-geodesy is a rapidly developing field of active research. It is recommended to design the reporting module so that it can be easily extended to report additional parameters. There is, for example, an indication (Lior, Ziv, and Madariaga, 2016) that reporting values from the displacement *and* velocity time-series, as it is computed by the Kalman filter, can be used in a scheme to estimate the P-wave *predominant period* as in Lockman and Allen, 2007, possibly without the saturation problems normally associated with large magnitude earthquakes. The final co-seismic displacements, after dynamic motions have stopped, can be used to model fault orientation and slip. Those can be used to derive Tsunami forecasts (Dragert et al., 2007; Hoechner, Babeyko, and Sobolev, 2008; Hoechner, Babeyko, and Sobolev, 2013; Melgar, Bock, and Crowell, 2012; Melgar and Bock, 2013; Hoshiba and Ozaki, 2014).

8 Earthquake Magnitude Scaling Relationships

8.1 Pd Magnitude Scaling and Error Propagation

The associator module then needs to apply the scaling relationships given in Crowell et al., 2013 to compute and update the magnitude and magnitude error estimates for the respective earthquake.

For each P_d

$$M_{P_d} = \frac{\log(P_d) + A + C\ \log(R)}{B} \tag{8.1}$$

with

$$A = 0.893, B = 0.562, C = 1.731$$

P_d is displacement in cm, R hypocentral distance in km and $\log()$ the base 10 logarithm.

The reported standard deviations, Σ_d (eq. 6.4), are propagated through the corresponding error propagation law to quantify the uncertainty in the magnitude estimate:

$$\sigma_{M_{P_d}} = \frac{d}{d\,P_d} M\,(P_d) \cdot \Sigma_d = \frac{1}{\ln(10)\,B\,P_d} \cdot \Sigma_d, \tag{8.2}$$

here $\ln()$ is the natural logarithm.

8.2 PGD Magnitude Scaling and Error Propagation

For each reported $\mathbf{PGD}$

$$M_{PGD} = \frac{\log(P_{PGD}) + D}{E - F\ \log(R)} \tag{8.3}$$

with

$$D = 5.013, E = 1.219, F = 0.178$$

Here also the reported standard deviations (eq. 6.6) are used to quantify magnitude uncertainty:

$$\sigma_{M_{P_{PGD}}} = \frac{dM(P_{PGD})}{dP_{PGD}} \cdot \Sigma_{PGD} = \frac{\Sigma_{PGD}}{P_{PGD} \cdot \ln(10) \cdot (E - F \log(R))} \tag{8.4}$$

For both scaling relationships P_d and PGD displacements are in units of centimeters, distance from the hypocenter R is in units of kilometers. Figures 6 and 7 show the respective relationships and may also be used to determine the respective lower limits of the seismo-geodetic system, assuming a resolution in the order of $1cm$. Crowell et al., 2016 modify the PGD relationship with data for the Nisqually earthquake, mainly to enable a depth-estimate to establish the earthquake hypocenter. This is currently beyond the scope of our project.

The implementation of magnitude updates within the associator/correlator will be discussed in a different context.

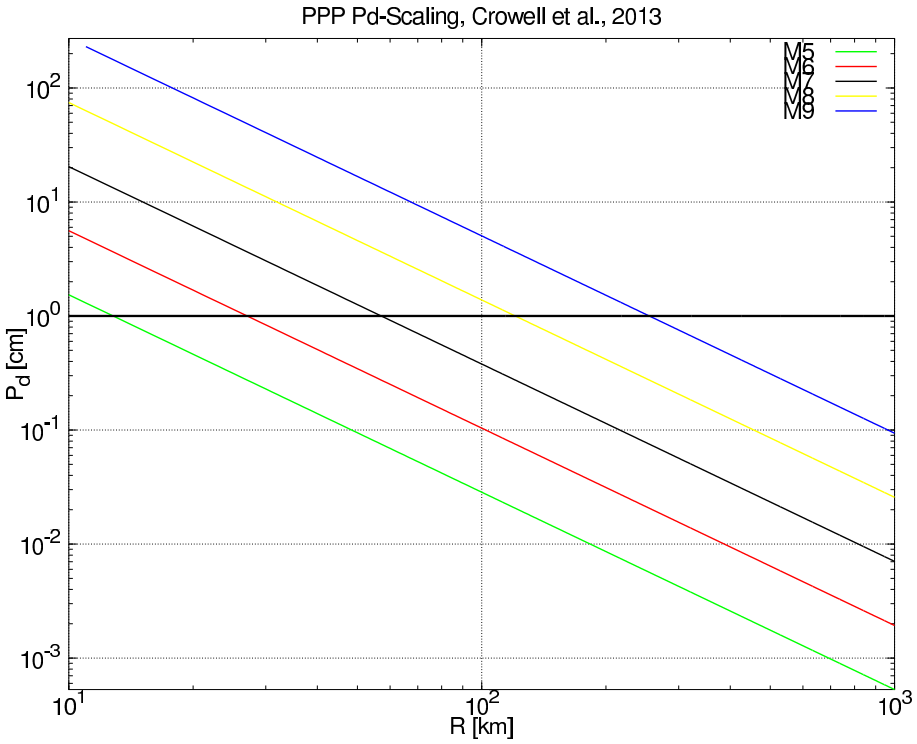

Figure 6: Seismo-geodetic P_d scaling relationships for mid-size to large earthquake magnitudes (after Crowell et al., 2013). Note, that for an M_w5 earthquake the limits of resolution of the seismo-geodetic system ($\approx 1cm$, thick black line) are reached at a hypo-central distance of about 10 km.

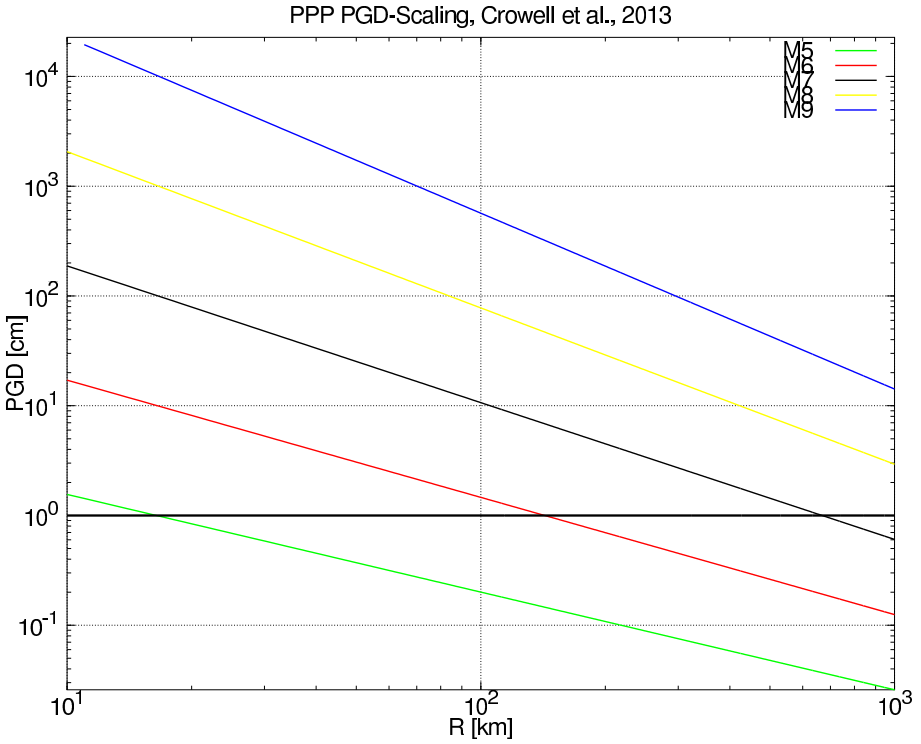

Figure 7: Seismo-geodetic PGD scaling relationships for mid-size to large earthquake magnitudes (aft. Crowell et al., 2013). The limits of resolution for the seismo-geodetic system ($\approx 1cm$, thick black line) are reached for M_w5 at a hypo-central distance of about 10 km, for M_w6 at about 100 km.

References

Aoi, S., T. Kunugi, and H. Fujiwara (2004). "Strong-motion seismograph network operated by NIED: K-NET and KiK-net". In: *Journal of Japan Association for Earthquake Engineering* 4.3, pp. 65–74.

Bar-Shalom, Yaakov, X. Rong Li, and Thiagalingam Kirubarajan (2001). *Estimation with Application to Tracking and Navigation*. Ed. by Yaakov Bar-Shalom. John Wiley & Sons, Ltd. Chap. 6, 271 ff.

Bock, Y., D. Melgar, and B. W. Crowell (2011). "Real-Time Strong-Motion Broadband Displacements from Collocated GPS and Accelerometers". In: *Bull. Seism. Soc. Am.* 101.6, pp. 2904–2925. DOI: 10.1785/0120110007.

Crowell, B. W., Y. Bock, and M. B. Squibb (Dec. 2009). "Demonstration of Earthquake Early Warning Using Total Displacement Waveforms from Real-time GPS Networks". In: *Seismol. Res. Lett.* 80.5, pp. 772–782. DOI: 10.1785/gssrl.80.5.772.

Crowell, Brendan W. et al. (2013). "Earthquake magnitude scaling using seismogeodetic data". In: *GRL* 40, 60896094. DOI: 10.1002/2013GL058391.

Crowell, Brendan W. et al. (2016). "Demonstration of the Cascadia G-FAST Geodetic Earthquake Early Warning System for the Nisqually, Washington, Earthquake". In: *SRL* 87, pp. 930–943. DOI: 10.1785/0220150255.

Dragert, H. et al. (Dec. 2007). "Towards Real-time Recognition of Near-Field Tsunamigenic Earthquakes". In: *AGU Fall Meeting Abstracts*, A1.

Emore, G. L. et al. (2007). "Recovering seismic displacements through combined use of 1-Hz GPS and strong-motion accelerometers," in: *BSSA* 97.2, pp. 357–378. DOI: 10.1785/0120060153.

Hoechner, A., A. Y. Babeyko, and S. V. Sobolev (2013). "Instant tsunami early warning based on real-time GPS – Tohoku 2011 case study". In: *Nat. Hazards Earth Syst. Sci.* 13, pp. 1285–1292. DOI: 10.5194/nhess-13-1285-2013.

Hoechner, Andreas, Andrey Y. Babeyko, and Stephan V. Sobolev (2008). "Enhanced GPS inversion technique applied to the 2004 Sumatra earthquake and tsunami". In: *Geophys. Res. Lett.* 35.L08310, pp. 1–5. DOI: 10.1029/2007GL033133.

Hoshiba, M. and T. Ozaki (2014). "Earthquake Early Warning and Tsunami Warning of the Japan Meteorological Agency, and Their Performance in the 2011 off the Pacific Coast of Tohoku Earthquake (Mw 9.0)". In: *Early Warning for Geological Disasters*. Ed. by F. Wenzel and J. Zschau. Advanced Technologies in Earth Sciences. Springer-Verlag Berlin Heidelberg, pp. 1–28. DOI: 10.1007/978-3-642-12233-0_1.

Kalman, R. E. (1960). "A New Approach to Linear Filtering and Prediction Problems". In: *Transactions of the ASME, Journal of Basic Engineering* 1, pp. 35–45. DOI: 10.1115/1.3662552.

Li, Xingxing (2015). "Real-time high-rate GNSS techniques for earthquake monitoring and early warning". PhD thesis. Technische Universität Berlin.

Lior, Itzhak, Alon Ziv, and Raul Madariaga (2016). "P-Wave Attenuation with Implications for Earthquake Early Warning". In: *BSSA* 106, pp. 13–22. DOI: 10.1785/0120150087.

Lockman, Andrew B. and Richard M. Allen (2007). "Magnitude-Period scaling relations for Japan and the Pacific Northwest: Implications for Earthquake Early Warning". In: *BSSA* 97.1B, pp. 140–150. DOI: 10.1785/0120040091.

Mayhew, David McNeil (1999). "Multi-rate Sensor Fusion for GPS Navigation Using Kalman Filtering". MA thesis. Virginia Polytechnic Institute and State University.

Melgar, D., Y. Bock, and B. W. Crowell (2012). "Real-time centroid moment tensor determination for large earthquakes from local and regional displacement records". In: *Geophys. J. Int.* 188, 703 – 718. DOI: 10.1111/j.1365-246X.2011.05297.x.

Melgar, Diego and Yehuda Bock (2013). "Near-field tsunami models with rapid earthquake source inversions from land- and ocean-based observations: The potential for forecast and warning". In: *Journal of Geophysical Research: Solid Earth* 118.11, pp. 5939–5955. ISSN: 2169-9356. DOI: 10.1002/2013JB010506. URL: http://dx.doi.org/10.1002/2013JB010506.

Melgar, Diego et al. (2013). "On robust and reliable automated baseline corrections for strong motion seismology". In: *JGR* 118, 11771187. DOI: 10.1002/jgrb.50135.

Niu, Jieming and Caijun Xu (2014). "Real-Time Assessment of the Broadband Coseismic Deformation of the 2011 Tohoku-Oki Earthquake Using an Adaptive Kalman Filter". In: *SRL* 85.4, pp. 836–843. DOI: 10.1785/0220130178.

Rosenberger, Andreas (2014). *Three component accelerometer signal processing for WARN*. Tech. rep. Ocean Networks Canada, University of Victoria.

Sagiya, T. (2004). "A Decade of GEONET: 1994-2003 - The continous GPS observation in Japan and its impact on eathquake studies". In: *Earth Planets Space* 56. DOI: 10.1186/BF03353077.

Saunders, Jessie K. et al. (2016). "Seismogeodesy Using GPS and Low-Cost MEMS Accelerometers: Perspectives for Earthquake Early Warning and Rapid Response". In: *BSSA* 106, pp. 1–21. DOI: 10.1785/0120160062.

Smyth, Andrew and Meiliang Wu (2007). "Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring". In: *Mechanical Systems and Signal Processing* 21, 706723. DOI: 10.1016/j.ymssp.2006.03.005.