User manual for `capuaf`: moment tensor inversion

Vipul Silwal, Celso Alvizuri, Carl Tape
Geophysical Institute, University of Alaska, Fairbanks, Alaska

Last compiled: July 14, 2022

Overview

This code is a modified version of the "cut-and-paste" (`cap`) code of *Zhao and Helmberger* (1994) and *Zhu and Helmberger* (1996). As requested in the `cap` documentation, any code that modifies `cap` should use a different name; hence, we refer to our version as `capuaf`. Table 1 lists publications with results obtained from `capuaf`. Moment tensor catalogs for these studies can be found in Scholarworks collections, as cited from within the published papers.

WARNING TO USER: This user manual is a work in progress and is probably not be up-to-date with the latest changes in the code.

REQUEST FOR FEEDBACK: If you detect bugs, have suggestions, or have questions, please post an issue on github. (Or email Carl Tape (`ctape@alaska.edu`), who will forward the request to `capuaf` developers.)

Table 1: Summary of studies using `capuaf`. DC = double couple moment tensors considered. FMT = full moment tensors considered. UNC = moment tensor uncertainties estimated. FMP = first-motion polarities used.

paper	DC	FMT	UNC	FMP	notes
Silwal and Tape (2016)	Y	–	Y	–	numerous comparison tests; confidence curve of *Tape and Tape* (2016)
Alvizuri and Tape (2016)	–	Y	(Y)	Y	misfit function plotted on lune
Alvizuri et al. (2018)	–	Y	Y	Y	confidence curve for full moment tensors; probability density functions for moment tensor source type (vw rectangle)
Silwal et al. (2018)	Y	–	–	Y	documentation of modified misfit function that includes polarities and also reward/penalty terms
Alvizuri and Tape (2018)	–	Y	Y	–	different static time shifts for Love and Rayleigh waves; interpretation in terms of crack-plus-double-couple model
Tape et al. (2015)	Y	–	–	Y	Minto Flats fault zone
Tape et al. (2017a)	Y	–	–	–	2000-02-03 Kaltag earthquake
Tape et al. (2018)	Y	–	–	–	2 very-low-frequency earthquakes, 4 earthquakes in Minto Flats fault zone; testing source duration and choice of bandpass

Contents

1 *Zhao and Helmberger* (1994)

1.1 Time shift

Cross-correlation function The relative time shift of the data f(t) from the synthetics g(t) can be estimated with the use of the cross-correlation function:

$$C(t) = \frac{\int_{-\infty}^{\infty} f(\tau)g(t+\tau)d\tau}{(\int_{-\infty}^{\infty} f^2(\tau)d\tau \int_{-\infty}^{\infty} g^2(\tau)d\tau)^{1/2}} \tag{1}$$

This is the normalized form of $C(t)$.

1.2 Seismic Moment

The moment is the ratio of the peak of the amplitude of the data to that of the synthetics

$$M_0 = \frac{Max(|f(t)|)}{Max(|g(t)|)} \tag{2}$$

1.3 Misfit

The error estimation in this earlier version of code involved both L1 and L2 norm.

$$L1 : ||f||_1^1 = \int_{t1}^{t2} |f(t)|dt \tag{3}$$

$$L2 : ||f||_2^2 = \int_{t1}^{t2} f^2(t)dt \tag{4}$$

in which [t1, t2] is the time interval, in which the seismogram is used. **The L1 norm emphasizes the high frequencies of the data, whereas the L2 emphasizes the low frequency**. Using Equation (2), the synthetics are defined as

$$d(t) = M_0 g(t) \tag{5}$$

The error is defined as (using either L1 or L2 norm)

$$e_{Lk} = \frac{||f-d||_k}{(||f||_k \times ||d||_k)^{1/2}} \tag{6}$$

Finally the error are implemented in the following manner:

$$e_1 = \frac{(e_{L1} + e_{L2} + (2e_{L1}^2 + 2e_{L2}^2)^{1/2})}{4} \tag{7}$$

For one station e_2 is defined the same way as e_1, except that the M_0 used in the equation 5 is the average moment $(1/n \sum_1^n M_0)$ of all the components used from the station. e_2 is a measure of the consistency of all components of one station, i.e. a minimum of e_2 gives closest ratios of the different components, say SH/P_nl of the synthetics to data. **In short, e_1 emphasizes the fit of the individual components, whereas e_2 emphasizes the consistency of all the components of one station**.

In the equation 7 equal weights have been given to the low frequency and high frequency components of the data.

The best solution is obtained by conducting the grid search in strike, dip and rake space, and finding the minimum of:

$$SOL : \min(E) = \frac{1}{n}\sum_{i=1}^{n}(e_{1i} + e_{2i}) \tag{8}$$

2 *Zhu and Helmberger* (1996)

The equations used in this paper are implemented in CAP. In this section, we copy text excerpts from *Zhu and Helmberger* (1996).

2.1 Misfit

We define an object function to measure the misfit error between u and s and search through the parameter space to find the global minimum of the object function. Misfit error is defined as the norm-k (L1 or L2) of the difference between u and s normalized by the norms of both u and s.

$$e = \frac{||u - s||_k}{||u||_k \times ||s||_k} \tag{9}$$

Because Pnl usually has smaller amplitude than surface waves, this normalization helps to weight Pnl and surface waves equally. It also prevents the inversion from being completely dominated by the strongest station, which is usually the nearest station, if several stations at different distance ranges are used. However, the amplitude information is lost during the normalization. Some of this information, such as amplitude ratios of Pnl-to-surface waves and SV-to-SH, provide important constraints on the source orientation and depth. A more severe problem with this normalization is that it introduces singularities in the source parameter space at those points where source orientation generates nodal synthetics (where the norm of synthetics vanishes). In the case when the data include nodal records, the grid search will miss the true minimum.

The misfit error using true amplitudes without normalization:

$$e = ||u - s||_k \tag{10}$$

2.2 Distance scaling factor

Using true-amplitude waveforms for source inversion usually leads to the problem of the closest station dominating the inversion when stations are distributed over a large distance range.

The misfits of surface waves have larger scatter than body waves, which are expected because surface waves are more easily affected by shallow heterogeneity. It has been shown that Pnl at a range of 300 to 1000 km is quite stable (Helmberger and Engen, 1980) and easily inverted for source mechanisms (Wallace and Helmberger, 1982). At closer range, the details of the Moho transition plays a more important role as well as the PL waves trapped in the shallow crust (Song and Helmberger, 1996). Since both of these features show strong local variation, we should expect the large scatter.

The misfit errors show a rapid decay with distance. Since radiation patterns have been taken out, this decay is related to the amplitude decay due to geometrical spreading and attenuation. To compensate for this decay, we introduce a distance range scaling factor and define the misfit error for a record at a distance r as:

$$\boxed{e = \left(\frac{r}{r_0}\right)^p ||u - s||_k} \tag{11}$$

There p is a scaling factor to give the record at r the same weight as that at reference distance $r0 = 100$. Any kind of norm-k **L1 or L2** can be chosen for calculating the data misfit.

If we assume a spherical geometrical spreading for body waves and cylindrical geometrical spreading for surface waves, an appropriate choice of p would be p=1 for body waves and p=0.5 for surface waves.

3 Implementation: CAP

The same misfit function as equation 11 is used and the **L2 norm** is chosen.

3.1 Distance weighting factor

First the distance weighting factor is computed

$$\mu = \left(\frac{r}{r_0}\right)^p \tag{12}$$

where r is the distance and $r_0 = 100$ km is the reference distance. Usually $p = 1$ for body waves and $p = 0.5$ for surface waves.

$$\begin{aligned} \mu_{pnl} &= \left(\frac{r}{r_0}\right)^1 & (13) \\ \mu_{surf} &= \left(\frac{r}{r_0}\right)^{0.5} & (14) \end{aligned}$$

At this point body and surface wave weight are imposed to evaluate the overall weighting factor

$$\begin{aligned} w_{pnl} &= (\mu_{pnl}) \times weight_{pnl} & (15) \\ w_{surf} &= (\mu_{surf}) \times weight_{surf} & (16) \end{aligned}$$

This weight is given in the input cap command and not the weight file.

We can examine how the parameters p and r_0 will influence the amplitudes of the waveforms used in the moment tensor inversions. Note that $\ln(A/B)$ provides a fractional difference bewteen positive quantities A and B.

$$\begin{aligned} A_{corr}(r) &= A_{raw}(r)\left(\frac{r}{r_0}\right)^p \\ \frac{A_{corr}(r)}{A_{raw}(r)} &= \left(\frac{r}{r_0}\right)^p \\ \ln\left(\frac{A_{corr}(r)}{A_{raw}(r)}\right) &= \ln\left(\frac{r}{r_0}\right)^p = p\ \ln\left(\frac{r}{r_0}\right) \end{aligned}$$

1. For distant stations ($r > r_0$):
   $p > 0$ will increase the amplitude
   $p < 0$ will decrease the amplitude

2. For nearby stations ($r < r_0$):
   $p > 0$ will decrease the amplitude
   $p < 0$ will increase the amplitude

3. In CAP $r_0 = 100$ km.

3.2 L2 norm of data and green's function

Before computing the norm, both the data and the green's function are subjected to the weighting factor:

$$\begin{aligned} u(t)_i &= w_i \times [f(t)_i]; \qquad i = [Pnl, Surf] & (17) \\ s(t)_i &= w_i \times [g(t)_i]; \qquad i = [Pnl, Surf] & (18) \end{aligned}$$

here f(t) is the recorded seismogram and g(t) is the green's function. Individual components of the data and green's function are picked and then multiplied with the suitable weighting factor depending it is the Pnl of Surface wave section. Each component (R,T,Z) of seismograms is chopped into Pnl and Surface wave section.

Example for a particular station:

$$
\begin{aligned}
[u(t)_{pnl}]_R &= w_{pnl} \times [f(t)_{pnl}]_Z && (19)\\
[u(t)_{pnl}]_Z &= w_{pnl} \times [f(t)_{pnl}]_Z && (20)\\
[u(t)_{surf}]_R &= w_{surf} \times [f(t)_{surf}]_R && (21)\\
[u(t)_{surf}]_Z &= w_{surf} \times [f(t)_{surf}]_Z && (22)\\
[u(t)_{surf}]_T &= w_{surf} \times [f(t)_{surf}]_T && (23)
\end{aligned}
$$

3.2.1 L2 norm

$$
\begin{aligned}
L2 : ||u||^2 &= \int_{t1}^{t2} u^2(t)dt && (24)\\
L2 : ||s||^2 &= \int_{t1}^{t2} s^2(t)dt && (25)
\end{aligned}
$$

For a vector signal, it is the inner product of the vector with itself.

$$
\begin{aligned}
L2 : ||\mathbf{u}||^2 &= \mathbf{u}^T\mathbf{u} && (26)\\
L2 : ||\mathbf{s}||^2 &= \mathbf{s}^T\mathbf{s} && (27)
\end{aligned}
$$

3.3 Cross Correlation

The normalized cross correlation function between u and s are computed using the same as Equation (1).

$$
\begin{aligned}
C(t) &= \frac{\int_{-\infty}^{\infty} u(\xi)s(\xi - t)d\xi}{(||u||^2\, ||s||^2)^{1/2}} && (28)\\
&= \frac{(u \star s)(t)}{(||u||^2\, ||s||^2)^{1/2}} && (29)
\end{aligned}
$$

This correlation function is the correlation between data and green's function. This is the normalized correlation function (or the **correlation coefficient** whose value is between 0 and 1. For find the actual actual correlation between two function (unnormalized):

$$
\begin{aligned}
C(t) &= (u \star s)(t) && (30)\\
&= \int_{-\infty}^{\infty} u(\xi)s(\xi - t)d\xi && (31)
\end{aligned}
$$

This is like convolution in some sense. To find the correlation of two functions, one function is slid over the another function. At some particular value of time τ_{max} this function $C(t)$ will have it maximum. This maximum correlation point is what we are interested inn matching the wave forms. And the shift τ_{max} is the relative shift between synthetics and data.

$$
\begin{aligned}
C_{max} = C(\tau_{max}) &= \max\left[corr(u(t), s(t))\right] && (32)\\
&= \max\left[(u \star s)(t)\right] && (33)\\
&= \max\left[\int_{-\infty}^{\infty} u(\xi)s(\xi - t)d\xi\right] && (34)\\
&= \int_{-\infty}^{\infty} u(\xi)s(\xi - \tau_{max})d\xi && (35)\\
&= \max\left[(u \star s)(\tau_{max})\right] && (36)
\end{aligned}
$$

Thus for the maximum correlation of two functions $u(t)$ and $s(t)$, we are multiplying the functions are shifting the synthetics by τ_{max} and then finding area under the curve $u(t)s(t + \tau_{max})$.

3.4 Moment magnitude

M_0, the is seismic moment $M_0 = \mu AD$, where μ is the shear modulus, A is the area of rupture and D is the average displacement. Relationship between moment magnitude M_w and seismic moment M_0 is given by *Kanamori* (1977) relation,:

$$
M_w = \frac{2}{3}(\log_{10} M_0) + k' \tag{37}
$$

Actual numbers are cited in Hanks and Kanamori 1979 GCMT catalog, *Aki and Richards* (1980), and *Shearer* (1999) uses the following equation:

$$
M_w = \frac{2}{3}(\log_{10} M_0) - \frac{2}{3}(16.1) \tag{38}
$$

However in the actual code, CAP uses the following equation to estimate the scaled seismic moment A from moment magnitude M_w :

$$
\begin{aligned}
M_w &= \frac{2}{3}(\log_{10}(A)) + \frac{2}{3}(20 - 16.1) && (39)\\
A &= 10^{(1.5M_w + 16.1 - 20)} && (40)
\end{aligned}
$$

To relate M_0 and A:

$$
\begin{aligned}
M_w &= \frac{2}{3}(\log_{10}(A)) + \frac{2}{3}(20 - 16.1) && (41)\\
&= \frac{2}{3}(\log_{10}(A) + 20) - \frac{2}{3}16.1 && (42)\\
\text{Therefore} & && (43)\\
(\log_{10}(A) + 20) &= \log_{10} M_0 && (44)\\
\log_{10}\frac{M_0}{A} &= 20 && (45)\\
\frac{M_0}{A} &= 10^{20} && (46)\\
A &= \frac{M_0}{10^{20}} && (47)
\end{aligned}
$$

P.S. Instead of M_0, A will be used to obtain the scaled moment tensor elements.

3.5 Source duration

Use -L flag. CAP now outputs the source function file (see OUTPUT_DIR/srcfile). The source function a trapezoidal function with rise_time = 0.5× source duration (Δt)

$$\Delta t = (int)10^{\frac{(Mw-5)}{2}+0.5} \tag{48}$$

For positive real numbers (int) function in perl is same as floor function of MATLAB, i.e., rounds towards zero.

Then it puts some sanity constrains so that duration do not exceed [1,9] seconds. Use -L flag for manual source duration length. See Figure 1 for variation of source duration with magnitude.

CAUTION: Using a wrong source duration can cause error in magnitude estimates. This become extremely crucial for small magnitude events. Smaller source duration (-L) will underestimate the magnitude and larger source duration will overestimate the magnitude.

```
$dura = 1 if $dura < 1;
$dura = 9 if $dura > 9;
```

3.6 Fault plane solution to Moment tensor

$$\begin{aligned}
M_{xx} &= -(\sin\theta\,\cos\sigma\,\sin 2\kappa + \sin 2\theta\,\sin\sigma\,\sin^2\kappa) && (49)\\
M_{yy} &= (\sin\theta\,\cos\sigma\,\sin 2\kappa - \sin 2\theta\,\sin\sigma\,\cos^2\kappa) && (50)\\
M_{zz} &= (\sin 2\theta\,\cos\sigma) && (51)\\
M_{xy} &= (\sin\theta\,\cos\sigma\,\sin 2\kappa + 0.5\sin 2\theta\,\sin\sigma\,\sin 2\kappa) && (52)\\
M_{xz} &= -(\cos\theta\,\cos\sigma\,\cos\kappa + \cos 2\theta\,\sin\sigma\,\sin\kappa) && (53)\\
M_{yz} &= -(\cos\theta\,\cos\sigma\,\sin\kappa - \cos 2\theta\,\sin\sigma\,\cos\kappa) && (54)
\end{aligned}$$

κ, strike of the fault

θ, dip angle of the fault plane

σ, slip angle

3.7 Green's function to Synthetics

When conducting the grid search over strike, dip and rake, the moment tensor is evaluated. Using this moment tensor and the azimuth of the station, the radiation pattern (**horizontal radiation coefficient**) are found. See *Jost and Herrmann* (1989).

$$\begin{aligned}
A1 &= -\frac{1}{2}(M_{xx} - M_{yy})\cos(2az) + M_{xy}\sin(2az) && (55)\\
A2 &= -M_{xz}\cos(az) - M_{yz}\sin(az) && (56)\\
A3 &= \frac{2M_{zz} - M_{yy} - M_{xx}}{6} && (57)\\
A4 &= -\frac{1}{2}(M_{xx} - M_{yy})\sin(2az) + M_{xy}\cos(2az) && (58)\\
A5 &= -M_{yz}\sin(az) + M_{xz}\sin(az) && (59)\\
A6 &= 0 && (60)\\
A7 &= \frac{M_{zz} + M_{yy} + M_{xx}}{3} && (61)\\
A8 &= 0 && (62)
\end{aligned}$$

To find the actual synthetics, these need to be multiplied by the scaled seismic moment A.
$\mathbf{A} = A * [A1\ A2\ A3\ A4\ A5\ A6\ A7\ A8]$.

The synthetics can be computed using the 9 green's functions and these 6 radiation coefficients:

$$\begin{aligned} d_z &= ZSS\ A1 + ZDS\ A2 + ZDD\ A3 + ZEP\ A7 && (63) \\ d_r &= RSS\ A1 + RDS\ A2 + RDD\ A3 + REP\ A7 && (64) \\ d_t &= TSS\ A4 + TDS\ A5 + TDD\ A6 + TEP\ A8 && (65) \end{aligned}$$

where R,T,Z are radial, transverse and vertical components of SS(vertical strike-slip), DS(vertical dip-slip) and DD(45^o dip-slip).

P.S. No Transverse components from vertical dip-slip fault and explosive source (TDD and TEP are 0)

WARNING: The above information may be inaccurate (?). For double-couple searches, there appear to be 5 unique coefficients and 8 Green's functions. For general moment tensor searches there appear to be 6 unique coefficients and 10 Green's functions. If these details are important to you, please double check the source code yourself to be sure. -Ryan

3.7.1 time shift

However instead of using the green's function the correlation function $C(t)$ is used. Correlation between the data and corresponding green's function. (i.e. instead of green's function ZSS, the correlation of ZSS and vertical component of data is used.

$$\begin{aligned} v_z &= corr(ZSS, u_z)\ A1 + corr(ZDS, u_z)\ A2 + corr(ZDD, u_z)\ A3 + corr(ZEP, u_z)\ A7 && (66) \\ v_r &= corr(RSS, u_r)\ A1 + corr(RDS, u_r)\ A2 + corr(RDD, u_r)\ A3 + corr(REP, u_r)\ A7 && (67) \\ v_t &= corr(TSS, u_t)\ A4 + corr(TDS, u_t)\ A5 + corr(TDD, u_t)\ A6 + corr(TEP, u_t)\ A8 && (68) \end{aligned}$$

where v_z, v_r, v_t are correlation of data and synthetics for a particular component. The value of time for which this correlation value comes out to be maximum, is the corresponding time_shift(τ) for that component.

Pnl wave (time_shift = $\tau 1$)

$$\begin{aligned} P_z &= corr(ZSS, u_z)\ A1 + corr(ZDS, u_z)\ A2 + corr(ZDD, u_z)\ A3 + corr(ZEP, u_z)\ A7 && (69) \\ P_r &= corr(RSS, u_r)\ A1 + corr(RDS, u_r)\ A2 + corr(RDD, u_r)\ A3 + corr(REP, u_r)\ A7 && (70) \\ P_{max} &= \max(w1 \cdot P_z + w2 \cdot P_r)_{\tau 1} && (71) \end{aligned}$$

Rayleigh wave (time_shift = $\tau 2$)

$$\begin{aligned} S_z &= corr(ZSS, u_z)\ A1 + corr(ZDS, u_z)\ A2 + corr(ZDD, u_z)\ A3 + corr(ZEP, u_z)\ A7 && (72) \\ S_r &= corr(RSS, u_r)\ A1 + corr(RDS, u_r)\ A2 + corr(RDD, u_r)\ A3 + corr(REP, u_r)\ A7 && (73) \\ S_{max} &= \max(w3 \cdot S_z + w4 \cdot S_r)_{\tau 2} && (74) \end{aligned}$$

SH wave (time_shift = $\tau 3$)

$$\begin{aligned} L_t &= corr(TSS, u_t)\ A4 + corr(TDS, u_t)\ A5 && (75) \\ L_{max} &= \max(w5 \cdot L_t)_{\tau 3} && (76) \end{aligned}$$

where w1 to w5 are the weights for particular component (as specified in the weight file).

3.7.2 New synthetics

After finding the point of maximum correlation, and shifting the synthetics, we obtain new synthetics of the form:

$$s'(t) = s(t - \tau) \tag{77}$$

The negative sign is used because positive time shift means, synthetics is earlier and needs to be shifted in positive t direction. In vector form :

$$\begin{aligned} \text{old synthetics :} \quad s(t) &= \mathbf{s} && (78)\\ \text{new synthetics :} \quad s'(t) &= \mathbf{s}' && (79)\\ \text{relationship :} \quad s'(t) &= s(t-\tau) && (80) \end{aligned}$$

3.8 L2 norm synthetics

Compute synthetics as in Equation (65) and find the L2 norm using the same formula as Equation (25).

$$\begin{aligned} s_z &= ZSS\ A1 + ZDS\ A2 + ZDD\ A3 + ZEP\ A7 && (81)\\ s_r &= RSS\ A1 + RDS\ A2 + RDD\ A3 + REP\ A7 && (82)\\ s_t &= TSS\ A4 + TDS\ A5 && (83) \end{aligned}$$

L2 norm

$$\begin{aligned} ||s||^2 &= \int_{t1}^{t2} s^2(t)dt && (84)\\ &= \int_{t1}^{t2} s'^2(t)dt && (85)\\ &= \int_{t1+\tau}^{t2+\tau} s^2(t-\tau)dt && (86)\\ &= (\mathbf{s}^T)(\mathbf{s}) && (87)\\ &= (\mathbf{s}'^T)(\mathbf{s}') && (88) \end{aligned}$$

3.9 Correlation between data and synthetics

The correlation between data $u(t)$ and synthetics $s'(t) = s(t-\tau)$ (new synthetics shifted by τ which gives maximum correlation), can be found by using Equation (36)

$$\begin{aligned} corr(u(t), s(t))_{max} = C_{max} = C(\tau_{max}) &= \int_{t1}^{t2} u(t)s(t-\tau_{max})dt && (89)\\ &= \int_{t1}^{t2} u(t)s'(t)dt && (90)\\ &= (u \star s)(\tau_{max}) && (91) \end{aligned}$$

3.10 Misfit

Finally the misfit e is given by:

$$\text{Integral form: } e = \int_{t1}^{t2} u^2(t)dt + \int_{t1}^{t2} s^2(t)dt - 2\int_{t1}^{t2} u(t)s(t-\tau_{max})dt \tag{92}$$

replacing $s(t)$ by shifted synthetics $s'(t)$

$$\begin{aligned} e &= \int_{t1}^{t2} u^2(t)dt + \int_{t1}^{t2} s'^2(t)dt - 2\int_{t1}^{t2} u(t)s'(t)dt && (93)\\ &= \int_{t1}^{t2} (u^2(t) + s'^2(t) - 2u(t)s'(t))dt && (94)\\ &= \int_{t1}^{t2} (u(t) - s'(t))^2 dt && (95)\\ \text{Vector form: } e &= (\underline{\mathbf{u}}^T)(\underline{\mathbf{u}}) + (\underline{\mathbf{s}}'^T)(\underline{\mathbf{s}}') - 2\,(\underline{\mathbf{u}}^T)(\underline{\mathbf{s}}') && (96)\\ &= (\underline{\mathbf{u}} - \underline{\mathbf{s}}')^T(\underline{\mathbf{u}} - \underline{\mathbf{s}}') && (97) \end{aligned}$$

This is the misfit error for a particular component of a particular station.

$u(t) \equiv \mathbf{u}$, is the recorded data (Pnl or Surface wave window)

$s(t) \equiv \mathbf{s}$, is the synthetic seismograms (Pnl or Surface wave window)

$s'(t-\tau) \equiv \mathbf{s}'$, is the synthetic seismograms shifted by τ to maximum correlation point (Pnl or Surface wave window)

For all N stations and all components:

$$\begin{aligned} E = \sum_{j=1}^{N}\sum_{i=1}^{5} e_{ij} &= \sum_{j=1}^{N}\sum_{i=1}^{5} w_{ij}(\underline{\mathbf{u}}_{ij} - \underline{\mathbf{s}}'_{ij})^T(\underline{\mathbf{u}}_{ij} - \underline{\mathbf{s}}'_{ij}) && (98)\\ &= \sum_{j=1}^{N}\sum_{i=1}^{5} (\underline{\mathbf{u}}_{ij} - \underline{\mathbf{s}}'_{ij})^T \mathbf{W}_{ij}(\underline{\mathbf{u}}_{ij} - \underline{\mathbf{s}}'_{ij}) && (99) \end{aligned}$$

where w_{ij} is the weight for i^{th} component at the j^{th} station. Since our data $\mathbf{u}_{ij}$ and synthetics $\mathbf{s}'_{ij}$ are vectors, $\mathbf{W}_{ij} = \mathbf{I}\,w_{ij}$. Dimensions of $\mathbf{I}$ are number of sample points in the corresponding time window.

The misfit is **not** normalized over number of stations or components.

This objective function E is minimized. The another way of looking at this is:

$$\begin{aligned} SOL : \min(E) &= \min(||u||^2 + ||s||^2 - 2\,corr(u(t), s(t))) && (100)\\ &= ||u||^2 + ||s||^2 - \max(2\,corr(u(t), s(t))) && (101)\\ &= ||u||^2 + ||s||^2 - 2\max[(u \star s)(t)] && (102)\\ &= ||u||^2 + ||s||^2 - 2[(u \star s)(\tau_{max})] && (103) \end{aligned}$$

3.11 Correlation percentage

Cross-correlation coefficient are computed using the normalized form. For a particular component, the correlation percentage is computed as follows:

$$\begin{aligned} cp_i &= 100 * \left(\frac{\max[corr(u(t), s(t))]}{\sqrt{||u||^2\,||s||^2}}\right) && (104)\\ &= 100 * \left(\frac{(u \star s)(\tau_{max})}{\sqrt{||u||^2\,||s||^2}}\right) && (105) \end{aligned}$$

could be simplified to:

$$cfg = 100 * \left(\frac{||u - s||^2}{\sqrt{||u||^2\,||s||^2}}\right) \qquad (106)$$

3.12 Variance reduction

Variance reduction is defined as improvement in the solution from some standard reference. In this case, the standard reference is when synthetics $s(t) = 0$ (no valid reason for assuming this). At $s(t) = 0$, the Variance Reduction $VR = 0$. And if our synthetics $s(t)$ perfectly matches the data $u(t)$, i.e. $s(t) = u(t)$, the $VR = 100$.

$$VR = 100 * \left(1 - \frac{E}{||u||^2}\right) \tag{107}$$

$$= 100 * \left(1 - \frac{||u - s||^2}{||u||^2}\right) \tag{108}$$

Dreger uses different Variance reduction formula,

$$VR = 100 * \left(1 - \frac{\sqrt{||u - s||^2}}{\sqrt{||u||^2}}\right) \tag{109}$$

For an acceptable solution, variance reduction is usually greater than 70%.

3.13 Depth Test

To obtain the best depth solution, inversion is performed at different depths and the minimum is found. To further get a better estimate, the minimum of the best fitting parabola is used.

However, for error-depth plot, a different measure is taken instead of misfit error, and then plotted against depth.

$$E\% = f \times \left(\frac{E_d}{E_{min}} - 1\right) \tag{110}$$

$E\%$, quantifying the error of depth d w.r.t. to error at best depth

f, degree of freedom where $f = n(N_{samp})$, where n is number of freedom per sample and N_{samp} is the total number of sampling points used for inversion (All stations body and surface waves).

E_d, error at depth d

E_{min}, error at best depth $E_{min} = \min(E_d)$

The default value of degree of freedom per sample $n = 0.01$. Changing this by order of 10, we can also control the steepness of the depth curve.

4 Changes made to CAP by UAF group

4.1 Changes in grid search

Uniform orientations provide a homogeneous distribution of double couple moment tensors. To achieve this, we use the coordinate $h = \cos\theta$ instead of dip angle θ (*Tape and Tape*, 2012).

θ_0, first element of search range in dip

$\Delta\theta_d$, search increment

N, number of samples to be generated

$\theta_1 = \theta_0 + (N - 1)\Delta\theta_d$, would be the final element of search range in dip

Then

$$\theta_i = \cos^{-1}\left(\cos(\theta_0) - i\left[\frac{\cos(\theta_0 - \cos\theta_1)}{N}\right]\right) \tag{111}$$

For the full moment tensor grid search, to obtain uniform spacing of points on the lune, we perform a similar transformation for the lune longitude δ:

$$\delta_i = \sin^{-1}\left(\sin(\delta_0) + i\left[\frac{\sin(\delta_1 - \sin\delta_0)}{N}\right]\right) \tag{112}$$

4.2 Changes in misfit function

1. Normalized the misfit error by number of components used. Error may now actually decrease by addition of 'good' stations. **normalization method**

$$||E_1(\mathbf{m})||_2^2 = \frac{1}{(P_{comp} + S_{comp})}\left[\sum_{i=1}^{P_{comp}} \frac{\int W_i||(\mathbf{u}_p)_i - (\mathbf{s}_p)_i||_2^2 dt}{(N_p)_i} + \sum_{i=1}^{S_{comp}} \frac{\int W_i||(\mathbf{u}_s)_i - (\mathbf{s}_s)_i||_2^2 dt}{(N_s)_i}\right] \tag{113}$$

$\mathbf{m}$, Model parameters (Strike κ,Dip θ,Rake σ)

u_p, observed P waveform

P_{comp}, number of P components used

$(N_p)_i$, number of sampling points for P (may vary from station to station)

u_s, observed S waveform

S_{comp}, number of S components used

$(N_s)_i$, number of sampling points for S (may vary from station to station)

W_i, weight given to that particular waveform

This equation can be simplified to following form:

$$||E_1(\mathbf{m})||_2^2 = \frac{\sum_{i=1}^{N_{comp}} (\mathbf{u}_i - \mathbf{s}_i)^t (\frac{W_i}{N_i})(\mathbf{u}_i - \mathbf{s}_i)}{N_{comp}} \tag{114}$$

where,

$\mathbf{u}_i$, observed waveform (body or surface wave)

$\mathbf{s}_i$, synthetic waveform (body or surface wave)

N_{comp}, number of components used (all body and surface waveforms)

N_i, number of sampling points for the waveform (may vary from station to station)

W_i, weight given to that particular waveform

Normalized misfit could be written as:

$$||E_r(\mathbf{m})||_2^2 = \frac{\sum_{i=1}^{N_{comp}} (\mathbf{u}_i - \mathbf{s}_i)^t (\frac{W_i}{N_i})(\mathbf{u}_i - \mathbf{s}_i)}{\sum_{i=1}^{N_{comp}} (\mathbf{u})^t (\frac{W_i}{N_i})(\mathbf{u})} \tag{115}$$

This is useful when we are comparing across different events. Other common names in the literature for this normalized or reduced misfit is 'reduced chi-square'. χ^2 could be converted to variance reduction (which have many different formulas in literature).

$$\begin{aligned} V.R. &= (1 - ||E_r(\mathbf{m})||_2^2) \times 100 && (116) \\ &= (1 - ||E_r(\mathbf{m})||_2) \times 100 && (117) \\ &= \ln\left(\frac{||E_1(\mathbf{m})||_2^2}{||d||_2^2}\right) \times 100 && (118) \\ &= \ln\left(\frac{||E_1(\mathbf{m})||_2}{||d||_2}\right) \times 100 && (119) \end{aligned}$$

Since variation reduction is only good for comparing

2. If higher weight is given to a good station whose waveform is matching well (higher correlation value), then overall misfit value reduces. vice versa, if higher weight is given to a bad station whose waveform is not matching well (low correlation value), then overall misfit value increases.

5 Creating posterior samples (*Silwal and Tape*, 2016)

Our misfit function needs to be modified because of two following reasons:

1. Low misfit value
   This basically is because of small amplitudes in displacement and velocity field of seismograms, hence also in L2 norm of data, synthetics and misfit. One common way of overcoming this is by normalizing it with data norm, but this leads to giving high weight in case of nodal stations (*Zhu and Helmberger*, 1996). Another way to get around this is dividing the misfit function by its minimum value and measuring its logarithmic variation.

$$\ln\left(\frac{E}{E_{min}}\right) \tag{120}$$

2. Another issue is with misfit function being too smooth. Possible reasons for this could be :
   (a) Not using the proper parameterization for moment tensor inversion (*Tape and Tape*, 2012).
   (b) Smooth misfit function also arises because of using exact magnitude M_w and not using full 6-parameter space for generating posterior samples. Our misfit function has only 3 model parameter (strike,dip,rake).

   To overcome this we scale the misfit by:

$$\frac{T}{\ln\left(\frac{E_{max}}{E_{min}}\right)} \tag{121}$$

Our final misfit function which is used for creating posterior pdf is:

$$E_2(\mathbf{m}) = E(\kappa, \theta, \sigma) = \frac{\ln\left(\frac{E}{E_{min}}\right)}{\ln\left(\frac{E_{max}}{E_{min}}\right)} \times T \tag{122}$$

where T is any multiplication factor. This T could be thought of as temperature in the energy function for *simulated annealing*.

5.1 Misfit function to posterior probability density function

See *Silwal and Tape* (2016); *Tape and Tape* (2016)

6 Combining polarity and waveform misfi (*Silwal et al.*, 2018)

Zahradnik et al. (2015) - We are NOT using the same approach. They used polarity as a constraint and searched for the possible solution
-X flag specifies the weight for combining polarity and waveform misfit.

For example, If the polarity weight is ω, i.e., -Xω

$$\begin{aligned}
polarity_weight &= \omega \\
waveform_weight &= 1 - \omega \\
polarity_error, \phi_{pol} &= \omega \times \frac{N_p}{N} \\
waveform_error, \phi_{wf} &= [SAME\ AS\ BEFORE] \\
total_misfit, \Phi &= \phi_{pol} + (1 - \omega) \cdot \phi_{wf}
\end{aligned}$$

See Figure 3 to see the effect of increasing the polarity weight on the obtained minimum total misfit solution.

Waveform error is not changed so that:

1. One can still use the post-processing scripts (MATLAB, perl). The waveform error and polarity misfit (number of stations at which polarity is not matching) are also saved in the binary files. If we make the change here then we have to make sure to make changes in other scripts.

Possible way of computing VR

Option 1 This didn't work out

$$\begin{aligned}
VR_{wf} &= 100 \times \left(1 - \left(\frac{\phi_{wf}}{u}\right)^2\right) \\
VR_{pol} &= 100 \times \left(1 - \left(\frac{N_p}{N}\right)^2\right) \\
VR &= \omega \cdot VR_{pol} + (1 - \omega) \cdot VR_{wf}
\end{aligned}$$

Option 2: This one is currently under testing and seems to work, and it is much cleaner too:

$$\begin{aligned}
total_misfit, \Phi &= \phi_{pol} + (1 - \omega) \cdot \phi_{wf} \\
VR &= 100. \times (1 - \Phi^2)
\end{aligned}$$

Here is how it looks in the code (sub_inversion.c):

waveform misfit

```
// waveform misfit
sol = get_tshift_corr_misfit(nda,obs0,max_shft,tie,norm,mtensor,amp,sol);
sol.wferr = sol.wferr/Ncomp;    // Ncomp = number of components.
sol.wferr = sol.wferr/data2;    // normalize by data

//---------------- combine polarity and waveform misfit---------------------------
// If -X flag is specified sol.err will contain the total misfit
 if ((int)pol_wt != 999){
```

```
     misfit_pol_weight = pol_wt; // this should come as an input from cap.pl (-X flag)
     misfit_wf_weight = 1 - misfit_pol_weight;
     sol.polerr = (float)misfit_pol_weight * misfit_fmp/nfm;
     sol.err = sol.polerr + misfit_wf_weight * sol.wferr;
     //fprintf(stderr,"---> %f %f %f %f\n",sol.err/data2, (float)misfit_fmp/nfm, total_misfit, mi
}
```

Implement station reward factor

Three reward factors are implemented:

```
// Implement station reward factor
// stn_rew = (float) (1.0 - ((2.0/pi)*atan(nda)))*5.0;
stn_rew = (exp(((float)-nda/7.0))*1.5)+0.5;
sol.err = stn_rew * sol.err;
```

Compute VR

```
// Compute VR
VR = 100.0 * (1 - (sol.err * sol.err));
```

TO DO: Add a figure

7 Adding reward factor (*Silwal et al.*, 2018)

1. longer time-window for Pnl and surface waves
2. broader bandpass for Pnl and surface waves
3. usage of more stations

7.1 Reward for longer time-windows and broader bandpass

The reward factor is applied alike to the body waves and the surface waves. Here is an example for the body wave:

Pw = length of Pnl window (example: 5 seconds)

Pband = Width of bandpass in Hz (example: 1- 10 Hz)

$$\boxed{Pnl\_reward = Pw \times Pband \qquad (123)}$$

This is how its implemented in cap.c:

```
---------------------Line 327 (cap.c)---------------------------
// Compute reward factors
  pnl_reward = (x1*(f2_pnl-f1_pnl));
  sw_reward = (y1*(f2_sw-f1_sw));

---------------------Line 602 (cap.c)---------------------------
// multiply weights by reward factors
// Add reward factor to each component
```

```
if (j<3) {
  spt->on_off = spt->on_off;
  spt->rew = sw_reward;
}
else {
  spt->on_off = spt->on_off;
  spt->rew = pnl_reward;
}

---------------------Line 700 (cap.c)---------------------------
        rec2 += spt->on_off*x2/(spt->npt * spt->rew);

---------------------Line 100 (sub_misfit.c)---------------------------
        x1 = x1/(spt->npt * spt->rew);
```

7.2 Reward when using more stations

While fitting the data with the synthetics we frequently encounter a case where the total normalized misfit when using more data is larger than the one obtained when using fewer data. This happens because it is much easier to fit a single waveform (or waveforms for a single/few stations) and could easily be done using a wrong MT solution (Figure 4).

Also the variance reduction (VR) is not an actual representation of the 'goodness' of the solution, since usually the VR decreases as we use more stations. In order to solve this issue we had to apply an exponential penalty function such that the fewer station inversions are penalized and the more data inversion is rewarded.

(Functions tested here are saved in: cap/test_station_reward.m)

$$k = (\exp(-N/CONST) * 1.5) + 0.5 \tag{124}$$

where k is the station penalty factor, N is the number of stations, and $CONST$ defines the shape of the penalty curve (see Figure 2). The shape of the curve ($CONST$) governs how severely the fewer station inversion needs to be penalized. An inverse exponential function scales between 0 and 1. For the test cases (Figure 4 and Figure 5) it was soon observed that scaling needs to be done in a wider range in order to reasonably scale the misfit values across wider range of stations. However, the number 1.5 in the above equation could possibly be event dependent and maybe removed by combining with the $CONST$. An additional constant 0.5 was added at the end to prevent the misfit from going to zero in case of a large number of stations.

Therefore the final scaled misfit is:

$$\Phi = k \times \Phi \tag{125}$$

Note: Other similar functions could also be used. I first started by using an inverse tangential function:

$$k = (1 - \frac{2}{\pi} * atan(N/CONST)) \tag{126}$$

But since this is just an ad-hoc function, using an inverse tangential makes things look more complicated.

References

Aki, K., and P. G. Richards (1980), *Quantitative Seismology, Theory and Methods*, W. H. Freeman, San Francisco, Calif., USA.

Alvizuri, C., and C. Tape (2016), Full moment tensors for small events ($M_{\mathrm{w}} < 3$) at Uturuncu volcano, Bolivia, *Geophys. J. Int.*, *206*, 1761–1783, doi:10.1093/gji/ggw247.

Alvizuri, C., and C. Tape (2018), Full moment tensor analysis of nuclear explosions in North Korea, *Seismol. Res. Lett.*, *89*(6), 2139–2151, doi:10.1785/0220180158.

Alvizuri, C., V. Silwal, L. Krischer, and C. Tape (2018), Estimation of full moment tensors, including uncertainties, for nuclear explosions, volcanic events, and earthquakes, *J. Geophys. Res. Solid Earth*, *123*, 5099–5119, doi:10.1029/2017JB015325.

Jost, M. L., and R. B. Herrmann (1989), A student's guide to and review of moment tensors, *Seismol. Res. Lett.*, *60*(2), 37–57.

Kanamori, H. (1977), The energy release in great earthquakes, *J. Geophys. Res.*, *82*, 2981–2987.

Shearer, P. M. (1999), *Introduction to Seismology*, Cambridge U. Press, Cambridge, UK.

Silwal, V., and C. Tape (2016), Seismic moment tensors and estimated uncertainties in southern Alaska, *J. Geophys. Res. Solid Earth*, *121*, 2772–2797, doi:10.1002/2015JB012588.

Silwal, V., C. Tape, and A. Lomax (2018), Crustal earthquakes in the Cook Inlet and Susitna region of southern Alaska, *Tectonophysics*, *745*, 245–263, doi:10.1016/j.tecto.2018.08.013.

Tape, C. (2016), Analysis of regional seismograms and 3D synthetic seismograms for the 2016-01-24 M_{w} 7.1 Iniskin earthquake in southern Alaska, ScholarWorks@UA at `http://hdl.handle.net/11122/6983` (last accessed 2016-11-03).

Tape, C., V. Silwal, C. Ji, L. Keyson, M. E. West, and N. Ruppert (2015), Transtensional tectonics of the Minto Flats fault zone and Nenana basin, central Alaska, *Bull. Seismol. Soc. Am.*, *105*(4), 2081–2100, doi:10.1785/0120150055.

Tape, C., A. Lomax, V. Silwal, J. D. Agnew, and B. Brettschneider (2017a), The 1904 M_{s}7.3 earthquake in central Alaska, *Bull. Seismol. Soc. Am.*, *107*(3), 1147–1174, doi:10.1785/0120160178.

Tape, C., V. Silwal, and S. Holtkamp (2017b), Step-response signals recorded during earthquakes in Alaska, ScholarWorks@UA at `http://hdl.handle.net/11122/7947` (last accessed 2019-12-07).

Tape, C., et al. (2018), Earthquake nucleation and fault slip complexity in the lower crust of central Alaska, *Nature Geoscience*, *11*, 536–541, doi:10.1038/s41561-018-0144-2.

Tape, W., and C. Tape (2012), A geometric setting for moment tensors, *Geophys. J. Int.*, *190*, 476–498, doi:10.1111/j.1365-246X.2012.05491.x.

Tape, W., and C. Tape (2016), A confidence parameter for seismic moment tensors, *Geophys. J. Int.*, *205*, 938–953, doi:10.1093/gji/ggw057.

Zahradnik, J., L. Fojtikova, J. Carvalho, L. V. Barros, E. Sokos, and J. Jansky (2015), Compromising polarity and waveform constraints in focal-mechanism solutions; the Mara Rosa 2010 Mw 4 central Brazil earthquake revisited, *Journal of South American Earth Sciences*, *63*, 323–333, doi:10.1016/j.jsames.2015.08.011.

Zhao, L.-S., and D. V. Helmberger (1994), Source estimation from broadband regional seismograms, *Bull. Seismol. Soc. Am.*, *84*(1), 91–104.

Zhu, L., and D. Helmberger (1996), Advancement in source estimation techniques using broadband regional seismograms, *Bull. Seismol. Soc. Am.*, *86*(5), 1634–1641.

A Time shifts

Layered (1D) models are often used as an approximation for Earth's heterogeneous structure for several reasons:

1. 1D Earth models allow for computational efficiency in thre form of rapid calculation of synthetic waveforms; this in turn allows uncertainties of model parameters to be throughly examined.
2. In some regions, the true Earth's structure is not far from 1D.
3. In some regions, the true/3D Earth's structure is not well known.

Using a 1D Earth model leads to challenges when trying to align (or 'time shift') synthetic waveforms with observed waveforms. Misalignment of waveforms—referred to as 'cycle skipping'—can lead to errors in estimating other source parameters, such as the moment tensor.

Time shifts are critical output information for a moment tensor inversion. We have explicity discussed time shifts in several studies (*Silwal and Tape*, 2016; *Alvizuri and Tape*, 2016; *Alvizuri et al.*, 2018; *Alvizuri and Tape*, 2018); these papers may provided a better starting point for this topic than this user manual.

Technical notes

There are three key categories related to time shifts:

1. measured time shifts (Table 2)
2. allowable time shifts (Table 3)
3. start and end times of time windows

In Tables 2 and 3 we try to adopt variable names that allow us to convey the key choices. In general, these names do not appear within the code itself (though perhaps in the future they should).

The onset time of the P wave (`Pobs`) is used to calculate time shifts. CAP looks in this order for (`Pobs`):

1. a value in a station's row-entry in the cap input weight file
2. the `A` SAC header in the data file
3. the SAC header on the Green's functions calculated from `fk`

These are three very different places to access `Pobs`, and it's important to note that some waveform extraction tools (e.g., the LLNL client database) may add headers unbeknownst to the user.

Table 2: Glossary of **measured time shifts** in `capuaf`. The tag of 'rel' refers to a relative time shift; all other time shifts are true/absolute.

name	formula	description	where found in capuaf	where displayed
`Pobs`		arrival time of observed P wave, relative to origin time in SAC file	weight file OR `A` header on SAC data file	
`Psyn`		arrival time of P wave calculated from a 1D model; this is done in fk when the 1D Green's functions are computed	sac header on fk Green's functions	
`dtP_pick`	`Psyn - Pobs`	travel time difference between Psyn and Pobs; based on picks from raw waveforms	e.g., -1.9 s (TUC)	beneath station name (when `Pobs` is used)
`dtP_CCrel`		cross correlation traveltime difference between shifted Psyn and Pobs		formerly displayed beneath P waveforms (consistent with `dtP_CC_rel_max`)
`dtP`	`dtP_pick + dtP_CCrel`	net time shift between Psyn and Pobs; based on cross-correlation of waveforms		beneath P waveforms; capmap spider plots; station scatterplots
`dtR_CCrel`		cross correlation between shifted-syn and obs Rayleigh wave		formerly beneath R waveforms
`dtL_CCrel`		cross correlation between shifted-syn and obs Love wave		formerly beneath L waveforms
`dtR`	`dtP_pick + dtR_CCrel`	actual time shift between syn and obs Rayleigh wave (on Z and R components)		beneath R waveforms
`dtL`	`dtP_pick + dtL_CCrel`	actual time shift between syn and obs Love wave (on T component)		beneath L waveforms
`Sobs`		arrival time of observed S wave, relative to origin time in SAC file; aka start time of Love/Rayleigh wave window (we have never used non-zero `Sobs`)	weight file	

Table 3: Glossary of **allowable time shifts** in `capuaf`. The tag on the variable is either 'min', 'max', or 'static'. The example numbers are for the case where a user wants allowable surface wave time shifts in the range of -2 s to +8 s, but also having an initial shift between waveforms (`dt_Ppick`) of -1.9 s. The net allowable time shift is then between -0.1 s and 9.9 s.

name	formula	description	where found in capuaf
`dtP_CC_rel_max`		allowed time shift for P, after syn has been shifted by `dtP_pick`	cap command line input
`dtR_static`		static time shift (default=0) so that min/max allowed time shifts will not be centered on 0	weight file (last column) e.g., 3 s
`dtR_CC_rel_max`		allowed time shift relative to `dtR_static`, after syn has been shifted by `dtP_pick`	cap command line input; e.g., 5 s
`dtR_min_rel`	`dtR_static - dtR_CC_rel_max`	minimum allowed surface wave time shift, after syn has been shifted by `dtP_pick`	not explicitly used; e.g., -2 s (= 3 - 5)
`dtR_max_rel`	`dtR_static + dtR_CC_rel_max`	maximum allowed surface wave time shift, after syn has been shifted by `dtP_pick`	not explicitly used; e.g., 8 s (= 3 + 5)
`dtR_min`	`dtR_min_rel - dtP_pick`	minimum allowed surface wave time shift	not explicitly used; e.g., -0.1 = -2 - -1.9
`dtR_max`	`dtR_max_rel - dtP_pick`	maximum allowed surface wave time shift	not explicitly used; e.g., 9.9 = 8 - -1.9

B Waveform selection criteria

Here we are trying to establish some standards for excluding waveform fits produced by CAP. The motivation is toward publication-quality figures. Time shifts play a major role in waveform selection, so please first read Appendix A.

1. Before doing any manual selection, make sure that

   (a) You have excluded the surface waves (and perhaps also body waves) for any stations that exhibit clipping. One straightforward assessment of clipping is to check if any of the three components have raw amplitudes that exceed a certain count level. From *Tape et al.* (2017b):

   > To assess clipping, we examine the raw, unfiltered waveforms and list the maximum of the absolute value counts on each channel for each station (e.g., Table **??**). For each station, the max value across all components is compared against a threshold value for the digitizer. We have assumed that all stations use a 24-bit digitizer except for GSN stations (COLA, KDAK), which use 26-bit digitizers. The threshold value is $q = \pm 2^{N-1}$, where N is the number of bits in the digitizer. We also list the number of time steps that a seismogram exceeds $0.8q$; this provides a quantification of square-like waves.

   The three components should be the unrotated components, which are written out to the RAW folder when running `pysep`. PROBABLY THIS STEP SHOULD BE PERFORMED AUTOMATICALLY IN PYSEP OR WHEN THE CAP INPUT FILE IS PREPARED.

   See also *Tape* (2016).

   (b) you have the right source duration (Figure 1); you may need to manually specify -L to obtain a shorter-than-default duration, especially for using body waves for small events

   (c) you are using L1 norm (this is the default)

   (d) the relative plotted amplitudes of body and surface waves are scaled correctly. This is just a plotting issue to make sure you can see the shapes of the waveforms. (But make sure you are looking at absolute amplitudes!)

   (e) **you have set the time shifts to the "correct" ranges**. This requires careful understanding of the options in Appendix A. **In the case of P waves, you may have to manually specify the onset time in the input weights file.** In the case of surface waves, you can specify a systematic time shift in the weights file if, for example, all time shifts appear to be systematically shifted from zero. This is equivalent to acknowledging that the structural model is uniformly slow (or fast) in all directions.

   Allowed time shifts depends on (examples given here are based on inversions in Alaska):

      i. Filter applied: For body waves (higher frequency content) you need to use a shorter time-shift compared to the surface wave. Example, 1–10 seconds body waves generally need $\sim$2 seconds of maximum shift; 16–50 seconds for surface waves can require upto 10 seconds of time-shift.

      ii. Distance range: Absolute value of time-shifts generally increases with epicentral distance (longer paths mean longer accumulation of time shift between 1D model and actual earth model). For example, time-shifts for surface wave (16–40 sec) for stations upto 100 km would be less than 5 seconds, but for stations at 400–500 km shifts could be around 10–12 seconds.

      iii. Source duration: In special cases, like VLFE (very low frequency earthquakes) which has much longer source duration (10-15 seconds for $M_{\rm w}$ 3.8 VLFE in interior Alaska as compared to 1 second duration for normal tectonic event of comparable magnitude) the time-shifts for surface wave (20–50 seconds) can go up to 18 seconds.

(f) your current moment tensor appears pretty close to correct

This requires having sufficient signal-to-noise levels within the chosen bandpass.

Note: If there is no reliable solution, then use the best-quality P polarities.

2. A published set of waveforms in a paper may be a subset of waveform fits. But the inversion should be done on a much larger set of waveforms that would be presented either in catalog results (e.g., ScholarWorks) or in a supplement. It's okay to have ugly fits in the "final" version for the catalog.

3. Here are our standards for catalog figures. **Keep in mind that these will be most effective when you are in the ballpark of the solution. This fine-tuning should only be done for the "final" solution.** Also keep in mind that we may choose to plot the green waveforms (`keepBad` in `cap_plt.pl`) or not. (It is useful for us to see the waveforms that we turn off, but it may be distracting to others.) These criteria are driven by amplitude anomalies, since our misfit function is amplitude-based and will be most affected by spurious amplitudes (even when using L1).

   PHASE 1 (getting in the ballpark of the solution):

   (a) Fix the magnitude and depth to the earthquake-catalog-listed magnitude and depth.

   (b) Use a "sensible" distance range for selecting stations. For example, for small magnitude event ($M_{\mathrm{w}} < 3$) one should set the maximum distance for selecting stations to 200 km. For intermediate magnitude ($M_{\mathrm{w}} > 3.5$) events one can go up to 500 km. This range might also vary depending on the crustal structure and topography. Also keep track of the available green's function (most are premade up to 500 km - `/store/wf/FK_synthetics/`)

   (c) Start by excluding the stations with the largest amplitude differences. (See also Point 1a above) Such differences may prevent you from viewing the waveform fits, since they are scaled by absolute amplitudes.

   (d) Exclude the set of stations at the largest epicentral distance above which stations all have poor SNR.

   (e) For some events with low SNR, you may be only able to use the vertical-component surface waves, since tilt tends to contaminate the horizontal components.

   PHASE 2:

   (a) Perform a magnitude search and depth search.

   Changing either one of these can have a major impact on the solution!

   (b) Consider adjusting time shifts, including P arrival times, and P polarities.

   (c) Exclude windows based on amplitude differences.

      i. Set weight zero for all **surface wave** waveforms whose amplitude differences are $|\mathrm{d}\ln A| \geq A_s$; these could be affecting the misfit function in a negative manner.

      - $A_s = 1.5$ for regional Alaska inversions

      ii. Set weight zero for all **body wave** waveforms whose amplitude differences are $|\mathrm{d}\ln A| \geq A_b$; these could be affecting the misfit function in a negative manner.

      - $A_b = 2.5$ for regional Alaska inversions (or for 2012 Nenana triggered event, which is full waveforms at the "body wave" periods)

Because surface waves are filtered at longer periods, which should be less sensitive to inaccuracies of the 1D velocity model, we have a stronger rejection criterion for fitting amplitudes (1.5 vs 2.5).

Do this incrementally: remove the largest amplitude anomalies, then rerun, etc.

(d) Set weight zero for all waveforms that max out the time shift. Or try to adjust the allowable time shift; see Point 1e above.

(e) If PV is set to weight zero, then also set PR weight to zero.

(Note: This means you can have PV only, but not PR only.)

(f) If, among the 3 surface wave windows, only Surf R is matching, then exclude this window.

(g) Ideally we would want to use both the body waves and surface waves, but there might be cases when either most of the body waves or surface waves need to be thrown out. In case you cannot use body waves at ∼90% of the stations, then perform a surface-wave-only inversion. Same goes for the surface waves.

Motivation: It looks odd to have a large set of body waves with a single surface wave (and vice versa).

(h) Time shifts for Love and Rayleigh waves should be "similar," since the propagation paths are sampling similar Earth structure (Love shallower than Rayleigh).

Use this to decide whether to turn off certain waveforms and whether to refine allowable time shifts. [NOTE THIS IS NOT A "RULE".] For a publication-quality solution, the time shifts should vary systematically as a function of azimuth. This can be detected from examination of the spider plot time shifts.

PHASE 3:

(a) Perform final magnitude and depth search.

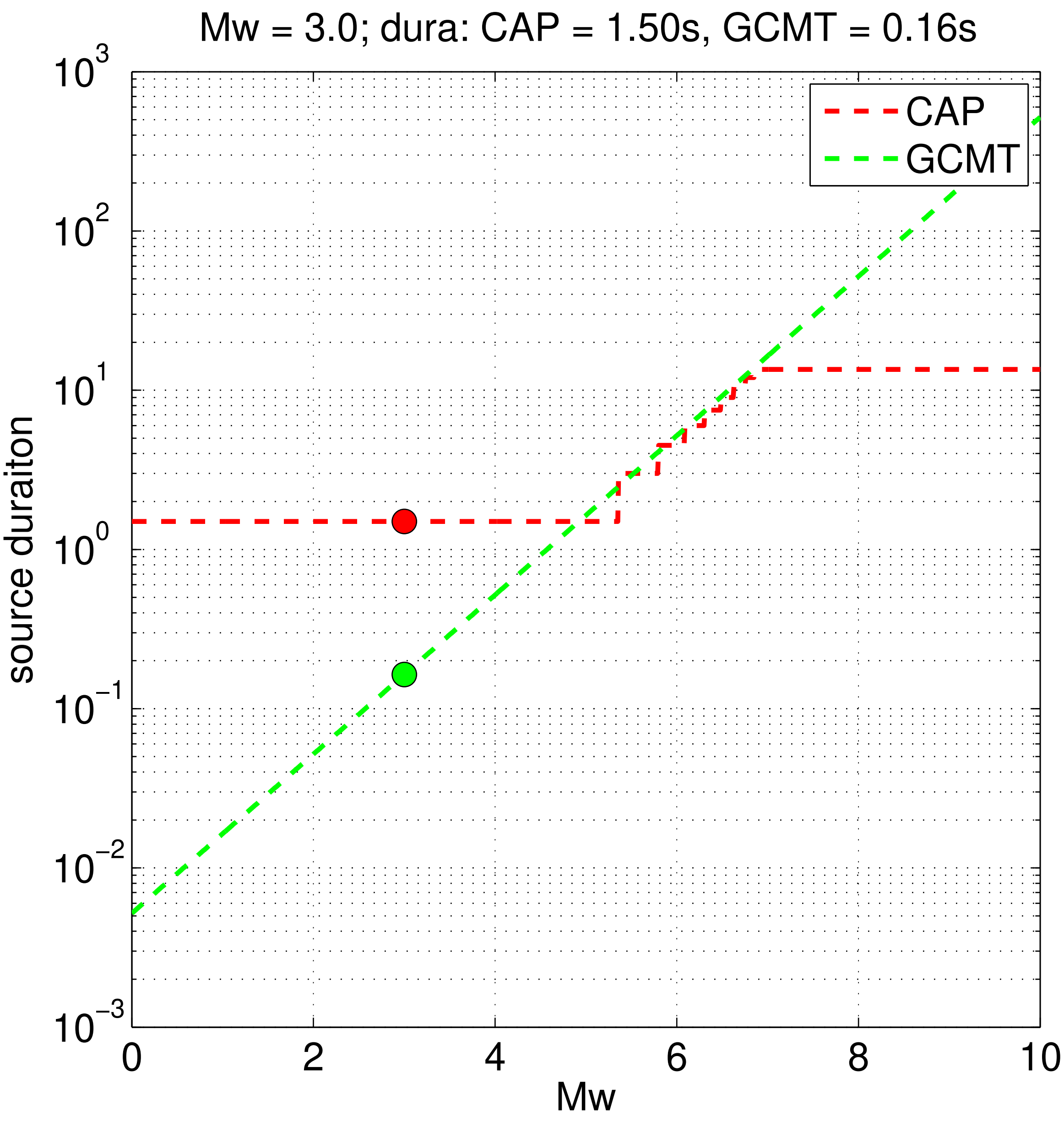

Figure 1: Source duration vs magnitude relation used in CAP (red) and by GCMT (green). Using a wrong source duration can make a big difference in the magnitude estimate. (ADD A FIGURE TO SHOW THIS).

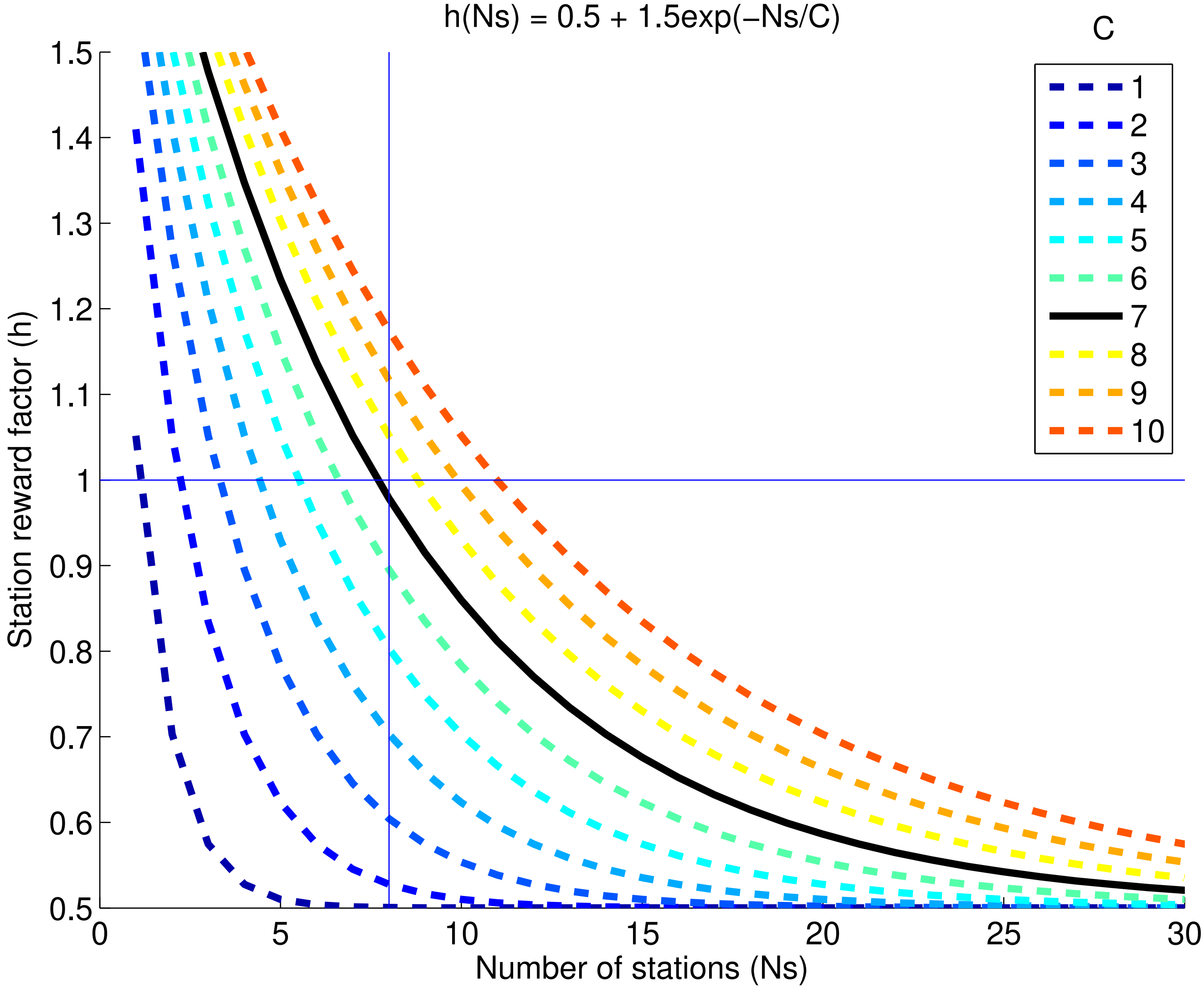

Figure 2: Station reward factor

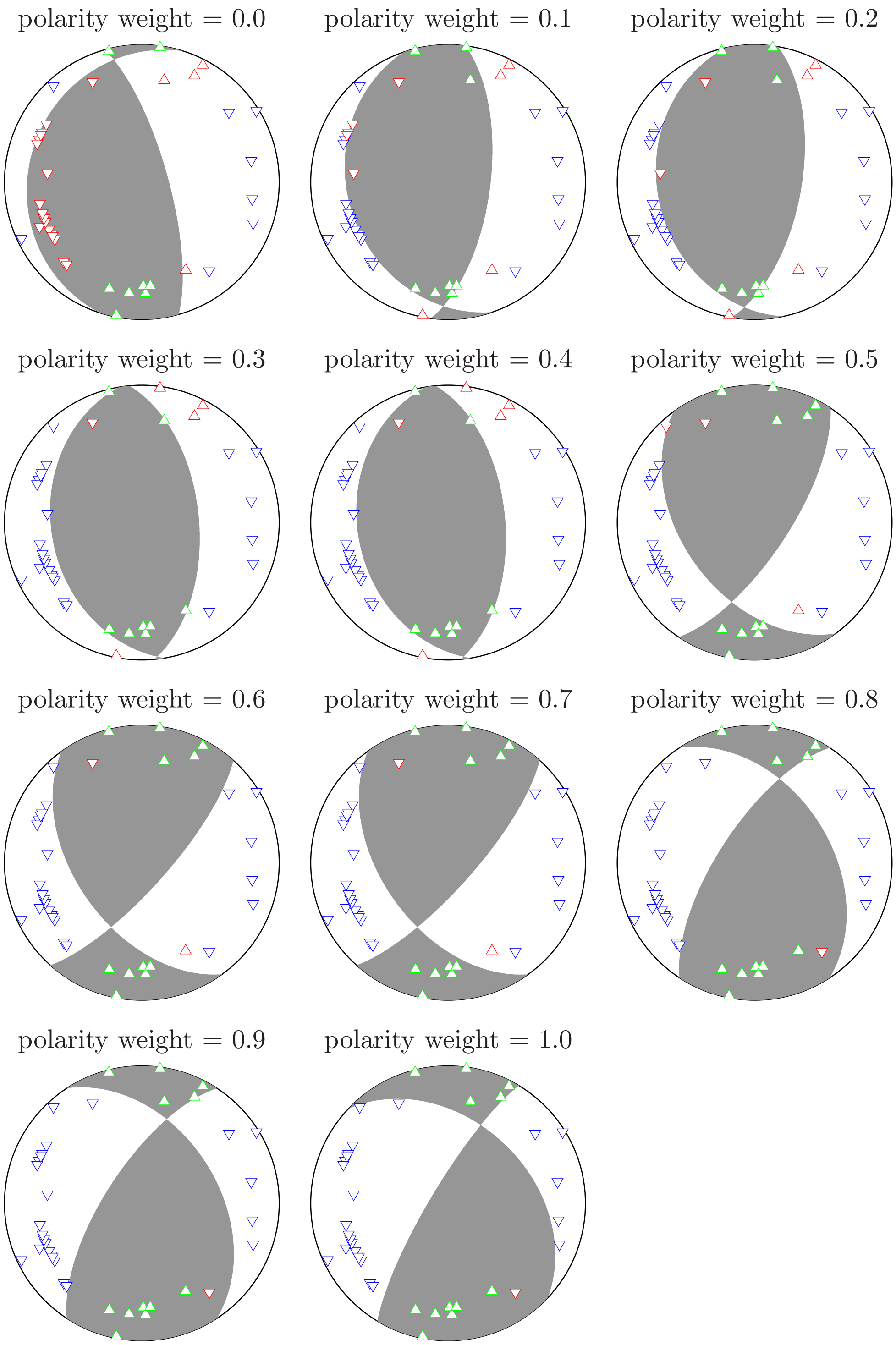

Figure 3: Effect of using different polarity weight factors. Marked in green are upward polarities, blue are downward polarities, and in red are mismatching polarities. Polarities are plotted at the lower-hemisphere piercing points on the beachball. Solution for polarity weight = 0.0 is obtained by minimizing waveform misfit only; solution for polarity weight = 1.0 is obtained using polarity only. Poorer the quality of inversion, more drastic are the variability in beachball to the changes in weight factor.

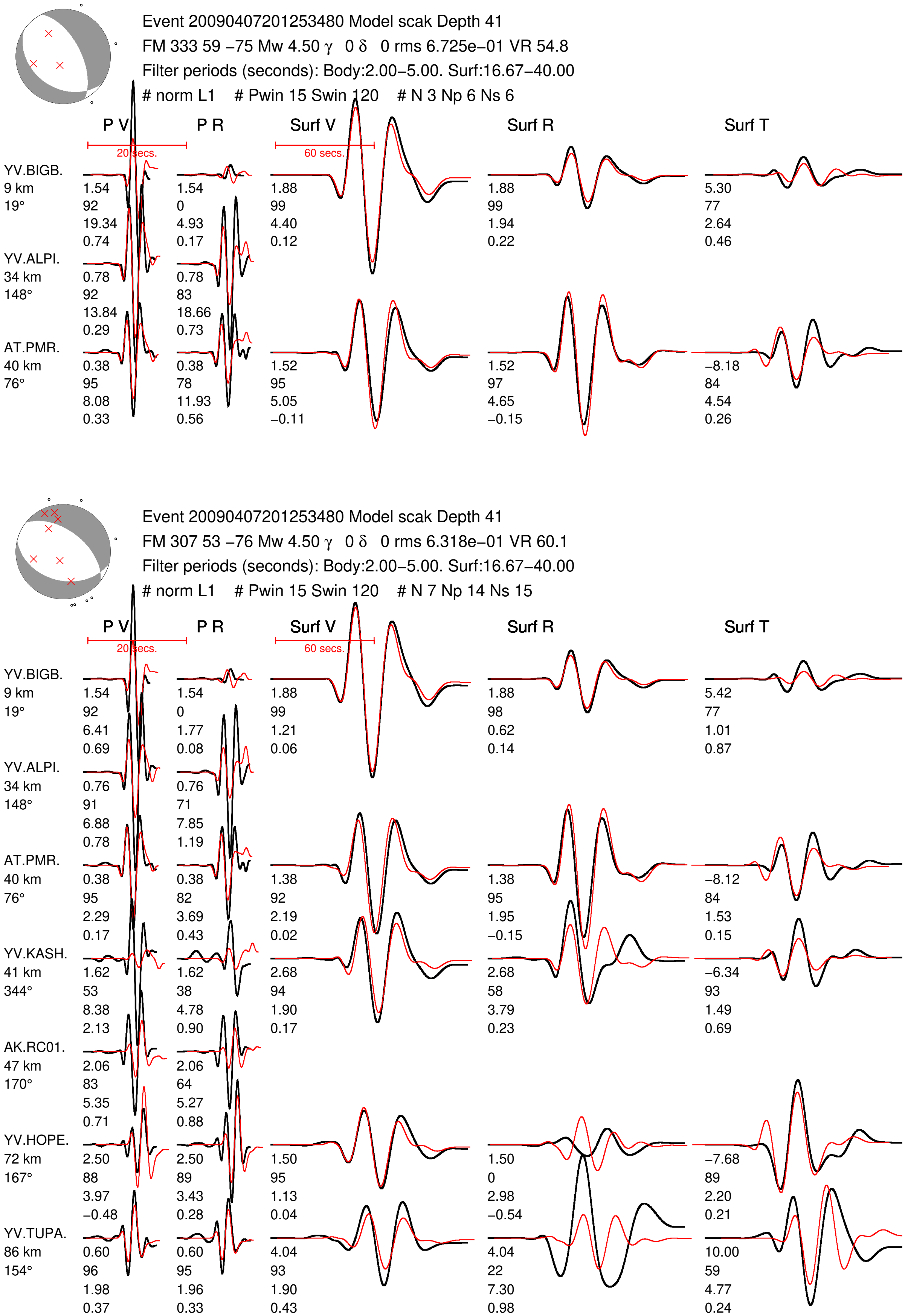

Figure 4: Waveform fits for CAP moment tensor solution when using 3 and 7 stations (top and bottom). Notice the increase in VR (also the decrease in the rms misfit) as we increase the number

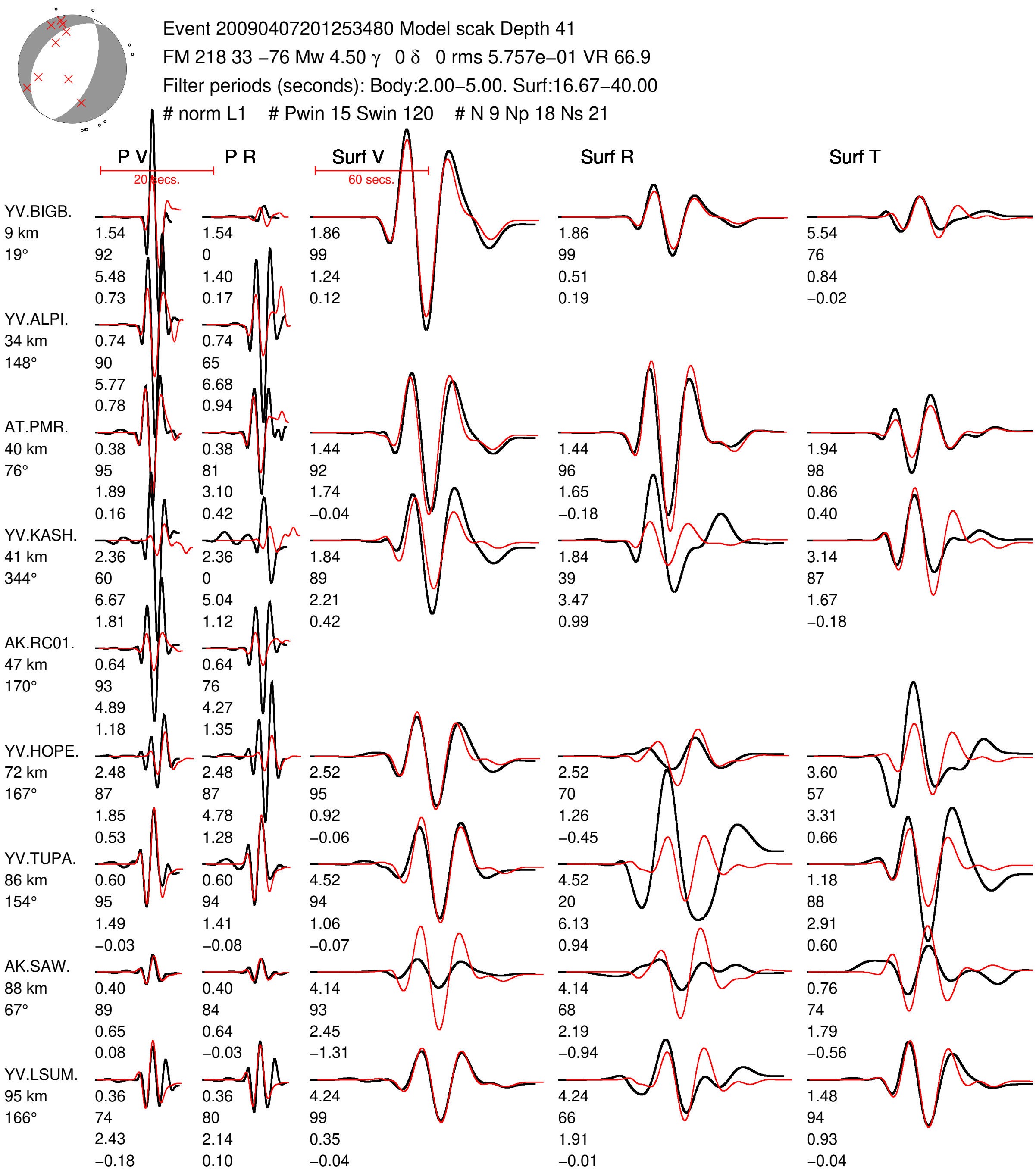

Figure 5: Waveform fits for the same solution as in Figure 5 when using 9 stations.