Travel time residuals and weighting:

Some things I've learnt

#1: Don't mistype them as
time travel residuals.
That's something else entirely.

What’s in a weight?

Linear location in SEISAN:

$$\sigma_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i)^2}$$

A ‘good’ hypocentre has RMS < 0.5, and individual station travel time residuals of a similarly low number.

If equally weighted, one bad pick can easily skew your hypocentre location, so need to introduce an unequal weighting term:

$$\sigma_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(w_i\, r_i)^2}$$

Within SEISAN, you can apply a weight at the time of picking:

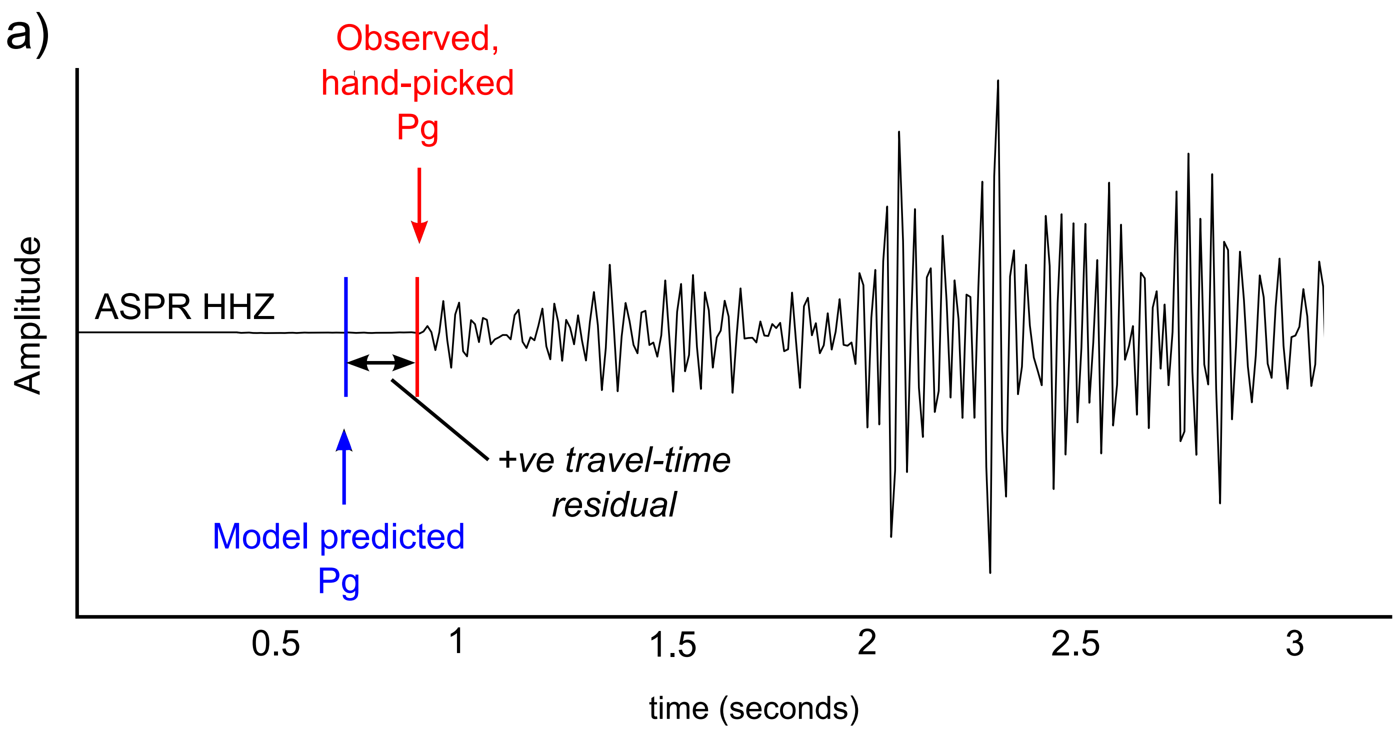
Assign weight between 0 and 4
(0 being highest quality, weighted 100%,
4 is the worst, weighted 0%)

However, this is not really a good ideas for a number of reasons:

- Very subjective
- Pick quality can change throughout the picking process, as you get more accustomed to picking
- Quality may unintentionally be considered for a single event, rather than over the dataset as a whole
- Doesn't give you a handle on the values required for NonLinLoc

Instead, an objective, statistical approach is preferred,
which relies on analysis of travel-time residuals

Single station, single earthquake



NEVER adjust a pick to lower the RMS, just to better fit the badly fitting velocity model.

Pick each phase as you see it….. S-phases/emergent are trickier

Sources of residuals:

1. **Errors in the applied velocity model.** Positive residuals indicate the calculated arrival times are too early compared with the observed values and on average the velocity model is too fast. The converse being true for negative residuals.

2. **Misidentification of phases.** This is particularly apparent for S-phases, which are more likely to be obscured by preceding P-phases and converted phases.

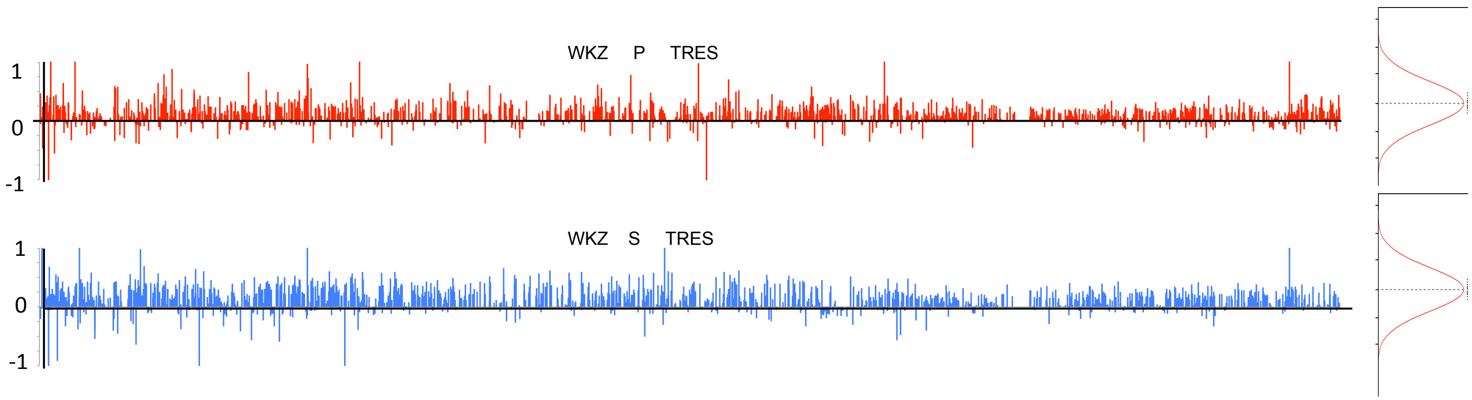
3. **Emergent versus impulsive arrivals**. Impulsive arrivals are sharp and the pick distribution associated with them is typically Gaussian (Buland, 1976). Emergent, less sharp arrivals are typically picked too late, and so their expected pick distribution is asymmetrical in time (Pavlis, 1986).

4. **Time stamp errors arising from GPS uncertainties**. Even assuming continuous, uninterrupted satellite communication, a phase error exists on the clock time recorded. This time error is typically of the order of a few microseconds.

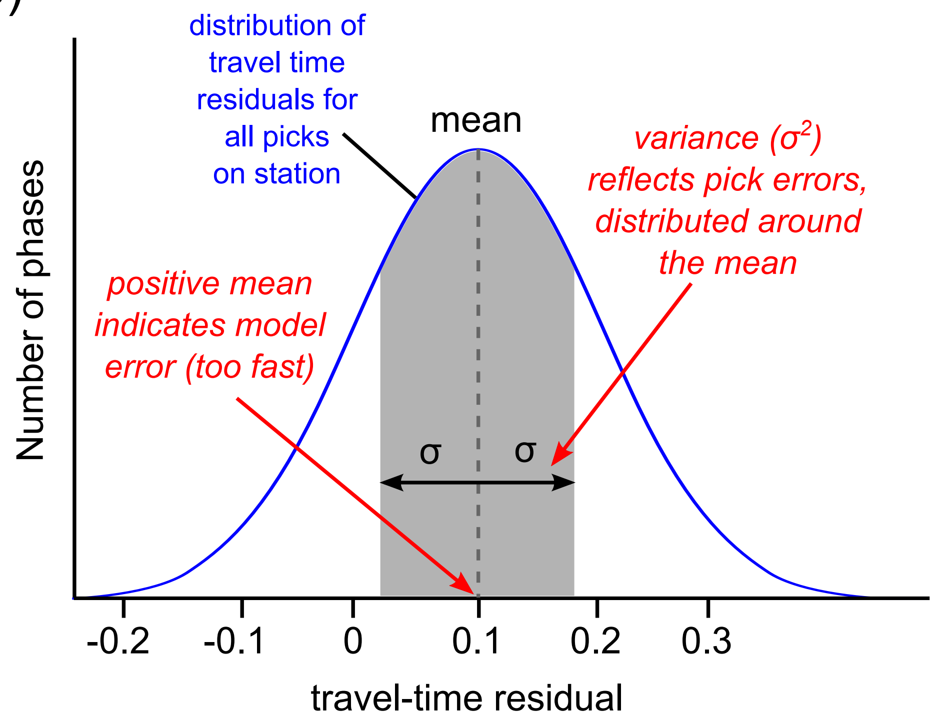
5. **Sampling limits**. All stations record at 100 Hz, meaning the lower limit on pick accuracy is ±0.005 seconds (Frémont and Malone, 1987).

6. **Reading accuracy**. Human error is a combination of some of the above factors, but is ultimately affected by the background noise of the data and how sharp the onset of the phase is. This effect can also vary with time as a user becomes accustomed to picking phases resulting in improved accuracy.

Single station,
Many earthquakes



b)



Gaussian distribution:

- Mean and standard deviation characterise your errors

1. **Errors in the applied velocity model.** Positive residuals indicate the calculated arrival times are too early compared with the observed values and on average the velocity model is too fast. The converse being true for negative residuals.

2. **Misidentification of phases.** This is particularly apparent for S-phases, which are more likely to be obscured by preceding P-phases and converted phases.

3. **Emergent versus impulsive arrivals**. Impulsive arrivals are sharp and the pick distribution associated with them is typically Gaussian (Buland, 1976). Emergent, less sharp arrivals are typically picked too late, and so their expected pick distribution is asymmetrical in time (Pavlis, 1986).

4. **Time stamp errors arising from GPS uncertainties**. Even assuming continuous, uninterrupted satellite communication, a phase error exists on the clock time recorded. This time error is typically of the order of a few microseconds.

5. **Sampling limits**. All stations record at 100 Hz, meaning the lower limit on pick accuracy is ±0.005 seconds (Frémont and Malone, 1987).

6. **Reading accuracy**. Human error is a combination of some of the above factors, but is ultimately affected by the background noise of the data and how sharp the onset of the phase is. This effect can also vary with time as a user becomes accustomed to picking phases resulting in improved accuracy.

The finer details of mean variations, arising from velocity model errors, form the basis of tomographic inversions

But also required in NonLinLoc:

LOCGAU - Gaussian Model Errors

required, non-repeatable

Syntax 1: **LOCGAU *SigmaTime CorrLen***

Specifies parameters for Gaussian modelisation-error covariances **Covariance** $_{ij}$ between stations **i** and **j** using the relation (Tarantola and Valette, 1982): **Covariance** $_{ij}$ = ***SigmaTime*** ^{2} **exp(-0.5(Dist** ^{2} $_{ij}$ **)/ *CorrLen*** ^{2} **)** where **Dist** is the distance in km between stations **i** and **j** .

SigmaTime (*float*, min:**0.0**) typical error in seconds for travel-time to one station due to model errors

CorrLen (*float*, min:**0.0**) correllaton length that controls covariance between stations (*i.e.* may be related to a characteristic scale length of the medium if variations on this scale are not included in the velocity model)

$$(C_T)_{IJ} = (\sigma^2)_{model} \exp\left(-\frac{D_{ij}^2}{2L_{corr}^2}\right)$$

(Tarantola and Valette, 1982):

σ_{model} is the mean model error (mean of Gaussian distribution) – averaged over all stations

L_{corr} is a length term related to the characteristic length scale of the velocity anomaly
Carolin played around with this and shows it's not very sensitive,
so set to the average station spacing.

Also need to know to objectively turn Gaussian residuals (arising from all the other errors) into a weight…

LOCQUAL2ERR - Quality to Error Mapping

required, *non-repeatable*, for phase/observation file formats that do not include time uncertainties ; *ignored*, *non-repeatable*, otherwise

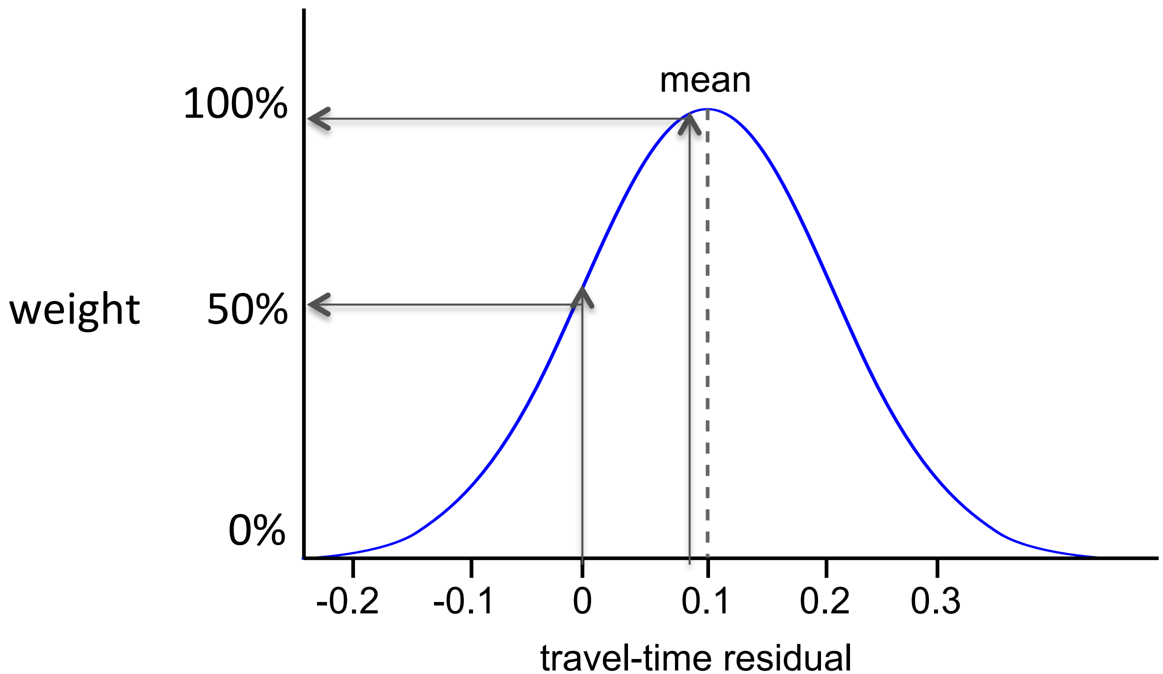
Syntax 1: **LOCQUAL2ERR** ***Err0*** **… … … …**

Specifies the mapping of phase pick qualities phase/observation file (*i.e.* **0**,**1**,**2**,**3** or **4**) to time uncertainties in seconds (*i.e.* **0.01** or **0.5**).

Err0 **…** ***ErrN*** (*float*, min:**0.0**) one time uncertainty value for each quality level that may be used in a phase/observation file. The first value *Err0* is assigned to picks with quality **0** , the second to picks with quality **1** , etc.

Notes:

1. NLLoc requires Gaussian timing error estimates in seconds for the data (phase picks), the **LOCQUAL2ERR** statement allows a conversion of commonly used integer quality codes to *float* time values.
2. Use a large, positive value (*i.e.* **99999.9**) to indicate a phase pick that should have zero weight (infinite uncertainty).



Jeffreys Weighting:
(Jeffreys, 1973)

Assumes the majority of earthquakes are well located, and a few outliers bias the travel-time residuals.

- First, exclude travel time residuals that fall outside 2 standard deviations of the mean as outliers and recalculated the mean (*v1*) and standard deviation (*σ1*) for the dataset excluding these outliers (iteration one).
- This process is repeated a second time (iteration two), again excluding outliers that fall outside two standard deviations of the mean of the dataset from iteration one.
- The remaining travel time residual distribution after two iterations had mean and standard deviations of *v2* and *σ2* respectively.

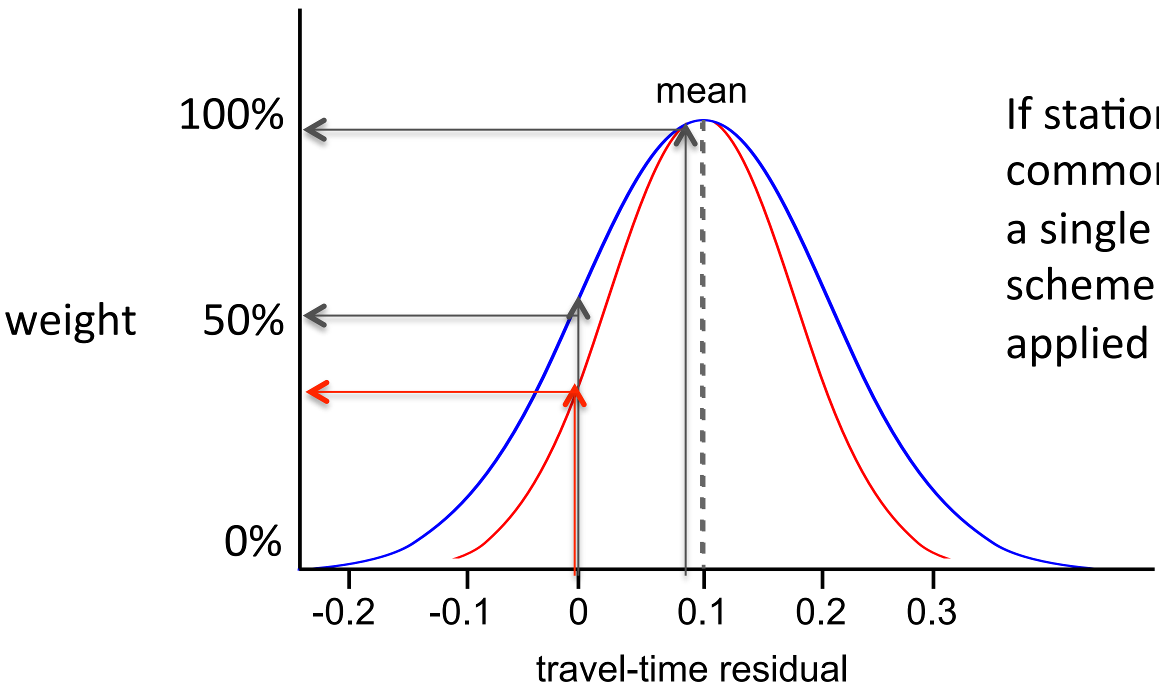
Then apply the following equation to each residual to calculate the Jeffreys' weighting.

$$w_r = \frac{1}{1 + \mu \exp\left(\frac{r^2}{2\sigma^2}\right)}$$

Where μ *is* the ratio of outliers to all values

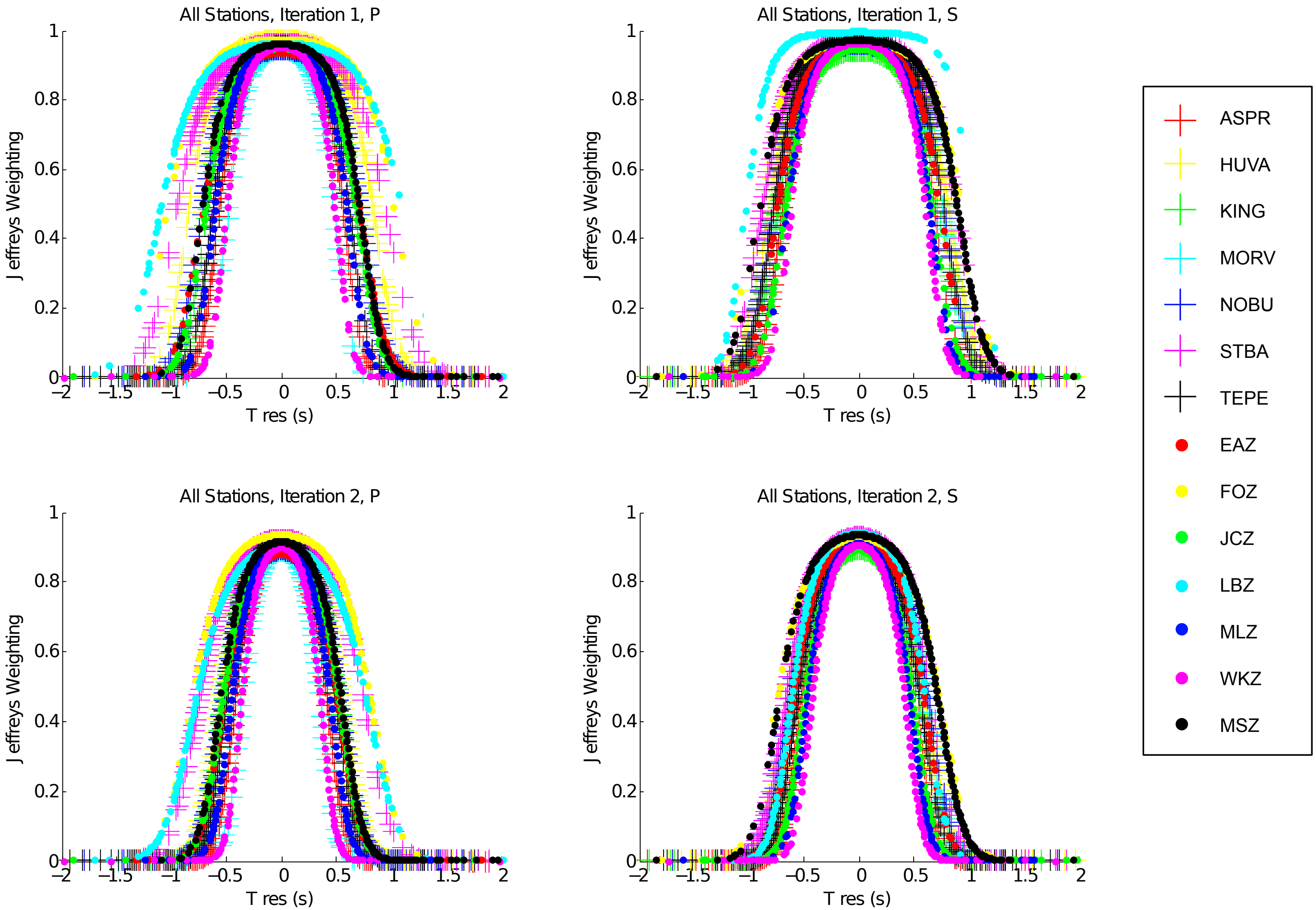
Many stations
Many earthquakes

Need to determine if Gaussian distributions are the same for all your stations…



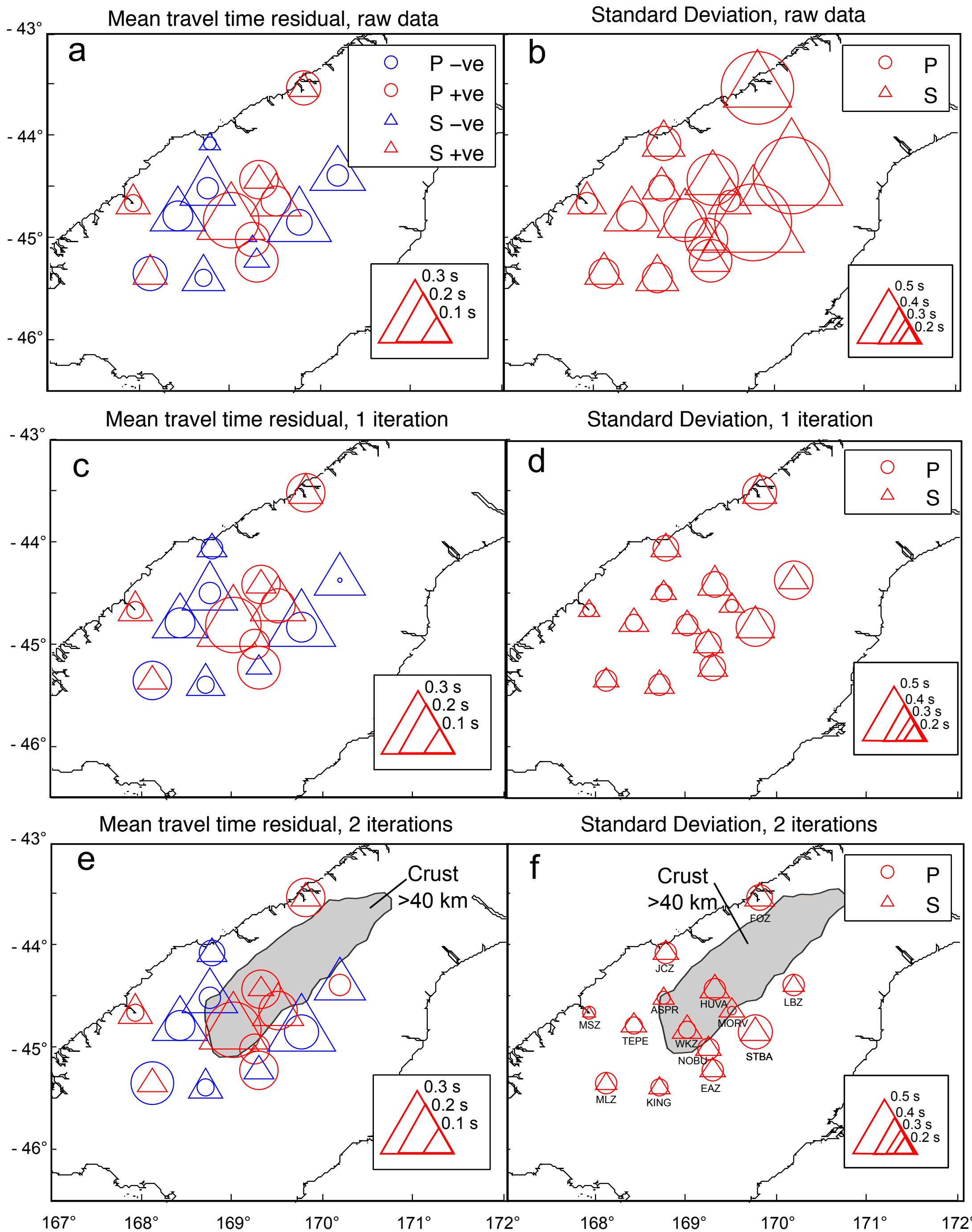
If stations don't have common variance, a single weighting scheme can't be easily applied

An example from COSA:



Positive means occur in the region of the thickest crustal root, and on moraine (FOZ). Model is too fast

Negative means generally distributed around root (overestimating crustal thickness, model is too slow at depth)



Largest variance in the east of the network

More scatter in picks coming from Fiordland, introduced from refracted Pn arrivals

Less clear after 2 iterations of removing outliers, but still present

Can also derive a distance weighting – weight closer stations more than farther stations

Use LOCGAU2 flags in NLLoc input file:

LOCGAU2 - Travel-Time Dependent Model Errors

optional, non-repeatable

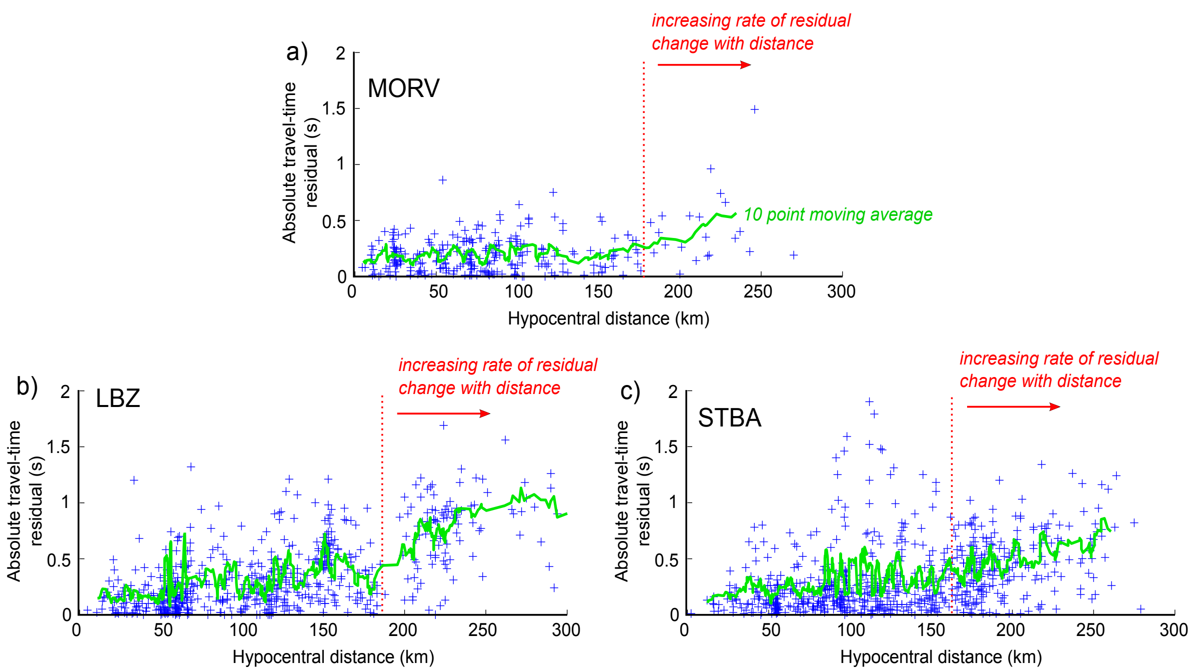
Syntax 1: **LOCGAU2** ***SigmaTfraction SigmaTmin SigmaTmax***

Specifies parameters for travel-time dependent modelisation-error. Sets the travel-time error in proportion to the travel-time, thus giving effectively a station-distance weighting, which was not included in the standard Tarantola and Valette formulation used by LOCGAU. This is important with velocity model errors, because nearby stations would usually have less absolute error than very far stations, and in general it is probably more correct that travel-time error is a percentage of the travel-time. Preliminary results using LOCGAU2 indicate that this way of setting travel-time errors gives visible improvement in hypocenter clustering. (can currently only be used with the EDT location methods)

SigmaTfraction (*float*, min:**0.0**, max:**1.0**) fraction of of travel-time to use as error

SigmaTmin (*float*, min:**0.0**) minimum trave-time error in seconds

SigmaTmax (*float*, min:**0.0**) maximum trave-time error in seconds



Can only be used if you have a linear increase in travel-time residual with distance

COSA data shows step-function, again indicating Pn arrivals from Fiordland events

In Conclusion:

- The problems I encountered will mostly likely arise for larger networks, where 1D velocity models are reasonably terrible.

- Smaller networks, like SAMBA, worked OK – not massive differences in variance.

- Can use a 3D velocity model – difficult to get working in NLLoc apparently, although I've never tried…

- The approach I ended up taking was just to assign a minimum error reading to each pick, effectively unweighted – which means your PDF scatter produced in NLLoc is effectively a minimum estimate of the error.

- NonLinLoc is better than linear location methods IF you understand the values you're plugging into it. Otherwise the process is pointless…

- Wibbly-Wobbly-Timey-Wimey-Stuff