Summary of points made by Langer et al, 2006.

Abstract:

* 5 transient seismic signal classes: VTE, REG, LPE, HYB, ROC.
* Identification and classification of these transients yield important information for the assessment of the state of the volcano system.
* Langer et al. applied ANN to 6000 events. Got 70% match to MVO classification. But for these ‘misclassified’ events, most of the them misclassified by MVO. So they manually revised the classifications. Retraining then reached 80%.
* Automatic classification was excellent in the identification of ROC and VTE events.
* Among misfits, erroneous attribution of LPE and HYB to ROC.
* Conclude that automatic classification with ANN is a powerful tool for handling large data masses and checking catalogs for consistent classification.

Introduction

* MVO established after the July 18, 1995 eruption, after 400 years of quiescence. Seismic monitoring has been re-established in 1992.
* Open-vent basaltic volcanoes such as Stromboli and Etna usually characterized by persistent volcanic tremor. Andesitic volcanoes like Soufriere Hills, seismic radiation is characterized by transient signals. Tremor at Soufriere Hills is often the superposition of repeated HYB and LP events.
* Different classes of signals with distinct waveform, frequency content and duration. The goal of data compression can be achieved by event classification.
* Refer to my paper for a definition of signals? Or a Montserrat paper? Or NMSOP?
* This scheme modified by MVO for Montserrat. VTE, REG, LPE, HYB, ROC. Gives a definition of each. HYB often recorded in swarms and sometimes merge to tremor. Cyclic swarms linked to cyclic tilt (Baptie et al, 2002), suggesting interconnection. Neuberg et al 2000 considers LPE and HYB as two end-members of a single event class, claiming a continuum exists between them (I have the data to check this claim).
* ROC events related to growth and collapse of lava dome (Calder et al, 2002), and can assume dimensions of PF. Usually 2-8 Hz bandwidth (check this), but significant energy can be present below 2 Hz (Luckett et al, 2002).
* Typically hundreds of transients detected per week. I should discuss the data reduction problem. ~700 MB per day I guessed at time. Continuous data. Detection. Classification. Location. This is what I was aiming for. In conclusions Langer says “cumbersome to handle. Data compressions and parameter extraction therefor necessary to exploit huge amount of information available for forecast and warning purposes”).
* Manual classification via visual inspection is slow, tedious, subjective, not done by trained seismologists, and even experts opinions are questionable/hard to reproduce. This is motivation for autoclassification.
* Description of supervised learning.
* Prior work Langer et al 2003 concluded ANN should work with a larger dataset (had only 336 records).

Data:

* Description of the digital seismic network, sensor types, sampling rate 75 Hz.
* They used data from 1996-1999 only, a period that reached its climax with devastating eruptions in 1997. I think I should expand this to 1996-2004 period for which I was there for initial paper. And then to 2005-2008 data, since it was a different seismic network. Then to both periods, comparing different models. And then to ASN data, since different data again and problems of clipping. And then broaden to all.
* Vertical component only.
* Durations 1-several minutes.
* They treated each trace as a separate event. This avoids complex reasoning about how to weigh significance of various stations, which could become obsolete as network changes. (Single channel classification, but ignoring trace IDs. I could try this. But I also want to try weighting. And there are some channels I would definitely want to weigh low anyway, e.g. MBMH, MBGH, MBRV).
* ~ 6000 records (records = traces; 6000 records ~1000 events since only verticals. I already have 522)

Application of the ANN:

* Description of ANN.
* ANN performance improves when the length of input data vectors are limited and phase alignment problems avoided (Falsaperla et al, 1996; Langer and Falsaperla, 2003).
* Following Langer et al, 2003, they use an information code which equalizes the length of input vectors, regardless of original signal duration.
* They use:
  + **autocorrelation function (acf)** which represents the spectral content using 16384 points for FFT, and then just use a fixed number of samples from it, since acf is always zero phase with maximum at time=0, phase alignment problems of raw waveforms eliminated,
  + statistical parameters such as sums of A, A^2, A^3, A^4 (statistical moments might be similar to energy, skew and kurtosis),
  + **amplitude ratio** between filtered 0.7-1.5 Hz and unfiltered signals,
  + **ratio of peak amplitude** versus RMS of 5-s before and 30-s after peak time

The statistical moments and amplitude ratios should distinguish brief peak transients like VTE from long duration RFE despite overlapping frequency content.

A thrust of my study will be that I combine metrics used by Marielle Malfante and Langer, plus some of my own. I should test autocorrelation traces – perhaps just add them as additional traces. And linearity, planarity and ellipticity traces when other components available. Then perform weighted reclassification. So for example, with 3 Integra stations and 5 Guralps, that is 18 channels in total. I could have: 18 waveforms, 18 auto-correlation functions, 5 linearity, 5 planarity and 5 ellipticity traces. This would be 51 traces in total. 51 separate models. Some models would suck. But this would help me figure out how badly they suck for classification. And it would use ALL THE DATA. And I could figure out ALL THE USEFUL FEATURES. And this would help design an efficient classification model to apply to the full dataset.

* Describe learning vs test groups, randomly.

Previous tests with a-priori classification

* They use the original classes assigned by MVO, which includes LPE+ROC.
* They removed events with peak-to-peak counts < 1000, to limit effects of noise. This left 4400 events. They randomly selected 3500 for the learning, and 900 for training (so about a 4:1 split). Success rate ~70%. No overfitting found.
* In a further test, they filtered events > 5000 counts. This left 2200 events, 1400 for learning, 800 for testing (so about a 2:1 split). And yet mismatch still around 30%. So success rate bounded by something intrinsic in the dataset (e.g. incorrect a-priori classifications – 4 examples shown in Fig 5) or methodology (different features needed?).

Reclassification with revised a-priori information

* Tanya revised classifications of 300 events (a partial revision only).
* This led them to examine largest 2400 events (p2p counts>5000) themselves. 200 events had poor signal quality. Here I should report on my methods for removing poor quality data (how many traces have I removed automatically?), and also on Ops Room procedures including my QC at the time.
* They relabelled LPE+ROC as ROC, so 6 to 5 classes. I could do this, and test whether it makes sense.
* They got 81% success rate in the training set, 78% in the test set. Can I beat this?
* Most ROC and VTE classified consistently (94.7% and 80.9% respectively, according to my calculations. But only 55.9%, 49.1% and 24.2% for LPE, HYB and REG. See “Langer2006results.xls”). Many LPE and HYB confused with ROC.

Discussion & Conclusions

* Continuous data acquisition brings problem of accumulating huge data masses. cumbersome to handle. Data compressions and parameter extraction therefor necessary to exploit huge amount of information available for forecast and warning purposes.
* Objective, reproducible.
* I think they kind of missed the original point that the original reason I approached them was it was impossible for MVO objectively classify the data. And I proposed 6 papers leading to automatic implementation.
* Note that their final results were obtained with p2p amplitude > 5000 counts. What thresholds have I used? Make figure.
* They claim evidence of separate between LPE and HYB, refuting Neuberg et al 2000’s claim.
* Scarcity of REG events in dataset is probably limiting factor.
* Luckett et al (2002) identified 3 classes of ROC (i) dominant frequency above 2 Hz, (ii) frequencies also 1-2 Hz, (iii) 1-2 Hz signal before rest of waveform. Seems we could try to train a model for each, if I can identify them, e.g. using spectrograms or filtered traces. Luckett speculates percentage of signal below 2 Hz linked to dome growth. Signal above 2 Hz may be due to material tumbling downhill, and overlaps with LPE and HYB signals in frequency. So there are intrinsic difficulties in separating signal sources.
* Recommends applying to whole Montserrat dataset. Yes, that was my original point. And would have happened had I not been pushed out of MVO.