doi: 10.1255/tosf.105

The 2010 Eyjafjellajökull volcanic ash sampling experiment: factors and conditions affecting field sampling variability

Kim H. Esbensen,[a] Pentti O. Minkkinen[b] and Hans S. Møller[c]
[a]Independent researcher, owner KHE Consulting, visiting, affiliated and guest professor at four universities
[b]Professor emeritus Lappeenranta University of Technology, Finland; owner SIRPEKA consulting
[c]Specialist; Bioenergy and Thermal Power COWI Consulting, Denmark

In August 2010, a party of Icelandic explorers undertook a sampling experiment of the famous Eyjafjella volcanic eruption local ash falls. The possibility of having 10 two-person sampling groups with very different Theory of Sampling (TOS) competence and experience proved an opportunity too interesting to miss. The Eyafjellajökull field experiment constitutes a Replication Experiment performed on a lot of significant generic complexity and heterogeneity, illustrating primary sampling when there is absolutely no control over the original lot and its history. In such cases, the effects on the total sampling variance (i.e. the total Measurement Uncertainty) reflects a compound of factors that cannot be fully resolved even after a comprehensive experimental design. The ultimate confounding concerns sampling competence vs lot heterogeneity. The lessons learned have easy-to-interpret educational relevance for many other types of material lots with similar characteristics and heterogeneity in science, technology and industry. It was also a lot of fun.

Introduction

The Icelandic volcano Eyjafjallajökull erupted in April and May 2010, causing one the largest impacts in the global awareness of volcanic eruptions in modern society. Very few, if any, living in northern Europe missed out on the detrimental effect of several total or partial shutdowns of the European airspace. The eruption partly also caused massive delays and flight cancellations on a global scale.

There is an abundance of information readily available on the internet for this volcanic eruption, eyewitness accounts, video documentations, travel accounts, scores of scientific studies and publications ("just Google it"). Should the reader of this feature be inclined to delve a little deeper into the geology, three references serve as a convenient portal (even though the last is in the Danish language, it boasts a series of magnificent illustrations which alone make its very affordable price worthwhile).[1–3]

Among the many spectacular effects from this eruption which, paradoxically, by normal standards of Icelandic volcanism was classified as a "minor, average eruption", was heavy local ash falls in the areas surrounding the Myrdalsjökull glacier under which the Eyjafjalla volcanic fissure is located, especially along the eastern and northern flanks.

Some three months after the eruption ended, a party of 20 Icelandic field trippers (actually a double 60-year birthday party) visited the Myrdalsjökull's northern-most glacial tongue Gigajökjull and, amongst other things, conducted a sampling experiment on this local ash fall. There were three experienced sampling experts in the company (a geologist, a chemist and an engineer, all members of what today is the International Pierre Gy Sampling Association, IPGSA), who conducted a crash course for all other participants on the principles of representative sampling of near-surface sedimentary deposits, e.g. soil and strata-bound sediments, including ash falls, Figures 1–3.

Experimental design

The experiment was designed to illustrate the effects of the most important factors affecting field sampling variability: varying training and experience (Theory of Sampling, TOS), grab sampling vs composite sampling and material heterogeneity—*temporal* (it is known that the general composition of the erupting lavas, and hence of

Figure 1. The April/May 2010 Eyjafjallajökull volcanic eruption produced an unusual abundance of volcanic ash. The finest particles produced the by now (in)famous atmospheric ash cloud that caused havoc to North Atlantic air traffic and beyond. Locally severe ash falls were an equally spectacular feature. Figure 1 shows the development of ash fall deposits, which constitute the basis for a 2010 field sampling exercise with which to characterise field sampling variability. Photo: J. Helgason / Shutterstock.com

Figure 2. The August 2010 Eyjafjallajökull sampling variance experiment. Ten two-person field groups with widely varying sampling competences each acquired a "field sample" for chemical analysis of local ash fall deposits close to the distal, northern Myrdalsjökull glacier tongue (Gigajökull), Iceland. Analyses of all samples are presented in Table 2 and in Figures 6 and 7. Figure 3 details on-site training of field crews and show details of the ash fall sampling conditions (see text).

the ash produced, changed its composition significantly during the *ca* three-month eruption) as well as *spatial* compositional heterogeneity. Ash fall may undergo a significant grain size differentiation during atmospheric deposition,[1–3] Figure 1.

One of the most interesting factors, of general sampling interest, concerned the possible inter-personal effect made manifest by organising 10 two-person sampling parties with very different experience and initial competence regarding the TOS. After proper TOS initiation, Figure 3, each team decided individually *where* and *how* to sample inside a 500 × 500 m area underlain by *apparently* relatively uniform ash fall deposits, at least from its superficial expression, Figure 1. Ash fall sediments consist of microscopic glass fragments, resulting from violent explosive gas-fragmentation of molten rock magma in conjunction with rapid quenching. All groups had been instructed about the possible effect caused by *aeolean* stratification (*wind-blown particle size segregation*) likely to have been operating during the ash fall(s), but to an unknown degree, emphasising the critical need for a *complete* depth section through the recognisable top ash fall deposit stratum, Figure 4, though for volcanological reasons it is not necessarily to be expected that there would be very large compositional differences in different ash particle size fractions due to a single eruption ash flow, which was the target for this experiment.

More importantly, sampling the deeper pre-eruption surface deposits, viz. vegetated topsoil or earlier volcanic deposits, was to be avoided at all costs (in a mining sampling context this horizon is termed the *sub-drill*). Thus, the target was identified as the *latest* (top-most) ash fall unit. While this is a relatively simple target to identify in the field, at least for geologists, it nevertheless constitutes a potentially significantly heterogeneous lot material with a complex deposition history for which sampling may not necessarily be a simple affair. At any rate, in mid-August 2010 the ash fall was there, the party was there, TOS was in the air—in short this was an opportunity too interesting to miss.

Upon completed instruction and extensive group discussions, all groups agreed that a single-scoop sample, a "grab sample", could not be accepted as representative sampling in view of the apparent heterogeneity manifestations of the target material at both micro- and meso-scales (but, see also below, because of other "samples" of ash falls from the same eruption, these undoubtedly with a grab sampling provenance). A concord was quickly reached in the field that *composite sampling* was to be employed by all groups. There was a certain spread in the opinions on *how many increments* would be needed under the prevailing circumstances. This was left at the discretion of the individual sampling groups, for reasons to be clear below, and the results presented below thus represent between three and ten increments.

A general misconception was underlined: sampling is *not* a simple mass-handling process. While the ultimate analytical test portion requires only 0.4 g of ash, it was emphatically pointed out, actually by one of the least experienced participants: "Surely this is not as easy as to grab *any* 0.4 g with a scoop of the appropriate volume—the ash is visibly non-uniform at this scale level. How could I possibly demonstrate that my single-scoop 'sample' is representative?" This statement was sweet music to an experienced sampling expert's ears.

This field sampling exercise does not constitute a particularly easy task, either for newly initiated samplers or for the sampling specialist (or even for the geologist involved) because of the largely

Figure 3. Field instruction on the principles of representative sampling (left). The inherent dangers of grab sampling were emphasised (right), as opposed to the virtues of composite sampling.

Figure 4. Differential local surface erosion interfering with sampling of the last Eyjafjella ash fall. Taking this effect into account, the right-hand photograph depicts sampling a complete depth section through the latest May ash fall layer, guided by the underlying pre-eruption vegetation manifestations (equivalent to the "sub-drill"). A completely (ideal) sample for a two-dimensional lot, such as a sedimentary layer, constitutes an increment in the form of a cylindrical drill core. The degree to which successive increments of the form illustrated deviate from this cylindrical form will induce various Increment Delineation Errors (IDE); there may also be vestiges of Increment Extraction Errors (IEE) when the spoon is in the hands of inexperienced samplers, i.e. spillage.

unknown deposition details. Therefore, it constitutes a fair educational context from which hopefully to elucidate the factors that contribute to primary sampling variability with clarity.

For comparison, on the internet, the University of Iceland, Geological Institute published *contemporaneous* ash analyses spanning the entire three-month period March–May 2010, Table 1, with which the ten replicated primary samples from the August 2010 experiment can be compared. It was also possible to obtain commercially available "Eyjafjallajökull volcanic ash", claimed to be "representative" of this famous eruption, a feature that was well reflected in the price asked for every 20 g sachet commercially available to the Iceland tourist in the latter part of 2010. Two of three procured sachets of this origin were also included in the present experiment.

The details of the field experiment can be appreciated in Figures 2–4, which illustrate the general sampling scene, the sampling conditions, the target lot and its meso-scale material features. All sampling teams were issued with identical sampling tools.

Chemical analysis

The ten field samples obtained, Figure 2, were subsequently subjected to identical secondary sampling (mass reduction) and sample preparation procedures, which took place at the Geological Survey of Denmark and Greenland (GEUS), Copenhagen. This was carried out by the same, very experienced sampler (present author) in order not to introduce unnecessary, compounding sub-sampling errors at these stages. The subsequent chemical analysis was carried out in an ISO-9001 accredited laboratory, certified to analyse combustion fly ash for major and trace compositions. For the purpose of displaying the above effects on the total sampling variability, the major element compositions may perhaps have sufficed alone, but trace elements were specifically included as these are generally much more effective in reflecting compositional heterogeneity. All chemical analyses reported here are courtesy of Vattenfall DK.

Since all secondary and tertiary mass reduction, sample preparation and analysis were scrupulously identical, adhering to TOS' stringent demands, differences between the individual analytical results presented below will only reflect meso-scale ash heterogeneity *and/or* the varying degree of competence/experience *between* these 10 groups as regards the primary field sampling. But how to differentiate between these two factors?

Every analytical result has a provenance

The final analytical results will always reflect the inherent *confounding* between effects originating because of the heterogeneous nature of the lot material (FSE, GSE) and/or because of whether the sampling procedure used is correct, or not (bias-free, or not) along the entire field-to-analysis pathway (ISE).[4] Due to material heterogeneity, analytical results will be dependent on the *specific* sampling procedure employed. An alternative sampling procedure will in general lead to different numerical analytical results, to the degree that sampling procedures are either representative or not, and will also impact on the empirical sampling variance. Whether such differences will be large or small is never known in advance. In the present case, serious attempts were made to avoid the basic bias-generating grab sampling procedures at the primary sampling stage, Figure 3, as well as in all laboratory procedures.[a]

---

[a] For reasons of keeping this article to a reasonable length, this is not the place also to present the basics of the Theory of Sampling (TOS), for which reason those seeking full explanations as to, for example, the so-called *correct sampling errors* (FSE, GSE), the complementary *incorrect*

Figure 5. Individual sampling groups in action.

Comparison, conditions and interpretations

As a means for comparison, the "official" rock and ash fall analyses published by the Earth Science Institute, Reykjavik University, will be used as benchmarks. As a group they in fact show a minor, but appreciable systematic difference to the field sampled samples in the present study, which is likely due to magma/lava compositional changes as the eruption progressed.[1,2]

But, more relevant to the present theme, there is no published information available as to the *methods* and *procedures* used for the sampling—*only* regarding analysis. There is nothing particularly unusual in this, however. The geological profession is, naturally, fiercely proud of their belief that: "… geologists *know* how to take a representative field sample", but this does not mean that TOS is irrelevant in the geological sciences. In fact, how to *assess* "representativity" based on non-complete sampling documentation, or a distinct lack thereof, often constitutes a moot issue in geology. All too often this issue is overlooked, or simply ignored, indeed it is often considered *irrelevant*: geologists train geologists, who train geologists, who train geologists… how to take a representative sample in the field. But *what if*… representativity is not a characteristic of the sample (of the analytical composition)? Such a thought is bordering on the heretical.

But, within TOS it is well known (this has been known for more than 60 years) that the adequacy and relevance of the analytical result *depends* on the specific sampling procedure used. It is fully possible to take a biased, non-representative sample (in TOS called a "specimen"), which is perhaps analysed with the ultimate precision, but the accuracy of which cannot be assessed. Within TOS it is well known that the qualifier "representative" can, and should, only be applied to the *sampling process*—not to the samples. It is not possible to ascertain the representativity status of a particular sample by its own characteristics—and this includes its analytical composition! Within this understanding, empirical sampling variability can arise both due to material heterogeneity *and/or* due to an inferior sampling process or even for other reasons (see footnote a above).

Staying with the traditional geological viewpoint, practice often devolves to a game of comparing total chemical analytical results at *face value*, i.e. as if analytical results always, universally are representative by *fiat* (when one does not know about sampling errors, there *are* no sampling errors). But the above argument explains why valid comparison is critically contingent upon full sampling documentation for all samples. Also: what is the status of a duly reported sampling procedure—representative, or not? It becomes clear that there is no such thing as a valid sampling comparison based on analytical results *alone*. This would be missing out completely regarding the provenance of the samples from which the analytical aliquot is but the last element.

Still, there is a perfectly feasible way out of this emerging dilemma: the TOS encompasses the complete set of ways and means needed in order to eliminate all bias-generating sampling errors (to ensure bias-free, accurate sampling) and further

---

sampling errors (ISE), *sampling bias* a.o., are referred to the readily available introductory literature, e.g. References 5 and 6.

how to reduce the effect of the remaining imprecision errors, i.e. reducing the sampling variability. For the reader who has just received a scientific jolt and wishes to know more, referral is made to the entire sampling literature… a portal to which can be found in References 6 and 7.

A field Replication Experiment

Thus, the present field experiment presents a TOS feature at the educational forefront. particularly relevant as regards “Replication Sampling Variability (RSV)”.[8] The Eyjafjella experiment serves well to illustrate the framework of this approach: what is the total *empirical* variability displayed by ten “replicate primary samples” of the same lot material? It is vital that the lot in question is a realistic lot, not some form of a made-up reference lot. This is so because all lots have both a spatial as well as a compositional heterogeneity aspect (perhaps even a differentiating temporal origin as well).

Table 1. Official Eayafjallajökull lava and ash fall compositions. Chemical analysis* of rocks and ash from the Eyjafjöll 2010 eruptions, Níels Óskarsson; major oxides in wt %, trace elements in mg kg^{-1} (ppm).

Available from
http://earthice.hi.is/eyjafjallajokull_2010_chemical_composition

Indeed, by not paying the necessary attention to the representativity of the specific sampling procedure, it is actually possible to sample in a fashion, which can be characterised as “precisely wrong” (*precise* because of a low spread of ten analytical results, high precision, but *inaccurate*, *wrong* (because of non-representative sampling). In general, a physical “average sample”, a *composite sample* consisting of a number of representative increments, is always to be preferred over *any* single grab sample. In this context, the present experiment aimed for ten such superior primary composite samples in order to leave only the inherent ash heterogeneity as the main factor behind the observable sampling variability. But in order to stay within a realistic sampling context (often primary samplers are employed with considerable different training, competence and experience), the present setup between sampler groups is fully realistic. This experimental setup will further information analogous to “reproducibility” in an analytical chemistry setting.

An alternative version of the RE could, for example, call for ten replicate primary samples executed by the *same sampler* (of course, also using the exact same sampling procedure), harking to “repeatability” in analytical chemistry. It is important to be

Table 2. Ash fall compositions from the 2010 Eyjafjallajökull field sampling experiment. The six first entries are from Table 1.

	SiO_2	Al_2O_3	Fe_2O_3	CaO	MgO	K_2O	Na_2O	TiO_2	MnO	P_2O	Ba	Co	Cr	Cu	Ni	V	Zn
SRG-2b	57.98	14.87	9.75	5.50	2.30	1.79	5.01	1.80	0.24	0.53	447	27.00	25.0	27.0	15.0	86.0	144
SRG-5a	56.73	14.65	9.93	6.11	3.15	1.65	5.04	1.88	0.24	0.43	425	33.00	97.0	35.0	41.0	114	137
BO1510	59.26	14.55	9.29	4.35	2.50	1.89	5.46	1.38	0.26	0.48	489	24.00	43.0	28.0	36.0	66.0	192
HO7510	58.24	14.89	9.17	4.70	3.17	1.78	5.18	1.50	0.24	0.48	449	29.00	57.0	29.0	55.0	89.0	161
PAA8510	59.02	14.43	9.38	4.20	3.16	1.86	5.22	1.40	0.26	0.48	478	28.00	60.0	26.0	78.0	78.0	165
GSV165-3	59.59	14.36	8.55	4.15	3.52	2.02	5.24	1.33	0.22	0.40	467	30.00	63.0	29.0	70.0	91.0	188
C1	61.40	15.00	8.41	4.55	2.86	2.13	4.99	1.31	0.20	0.20	421	27.70	78.8	22.2	49.5	85.0	142
C2	61.00	14.90	8.50	4.54	3.02	2.15	4.93	1.32	0.20	0.20	412	28.30	79.7	22.2	52.8	85.1	140
C4	61.10	14.90	8.51	4.58	2.95	2.12	4.97	1.35	0.20	0.21	419	29.40	81.4	22.6	54.7	87.9	148
C5	60.70	14.80	8.57	4.57	3.17	2.12	4.84	1.30	0.20	0.20	415	30.20	89.3	22.5	63.4	87.7	145
C6	61.40	14.90	8.39	4.46	2.84	2.15	5.00	1.30	0.20	0.20	411	28.30	74.1	22.5	51.9	85.3	148
C7	60.70	14.90	8.63	4.64	3.02	2.09	4.92	1.40	0.20	0.22	404	31.30	75.8	24.9	53.0	99.6	147
C8	60.20	14.70	8.73	4.68	3.43	2.24	4.79	1.33	0.20	0.20	409	31.40	92.6	22.0	69.8	87.3	145
C9	61.00	14.90	8.46	4.56	2.90	2.13	4.96	1.32	0.20	0.21	426	30.00	84.3	22.9	56.8	88.9	150
C10	59.90	15.00	8.71	4.96	3.23	2.01	4.80	1.44	0.20	0.22	403	33.20	97.4	27.0	61.5	101	145
C3-1	61.40	15.00	8.30	4.34	2.71	2.19	5.04	1.28	0.19	0.20	417	28.10	71.2	22.2	53.0	84.3	143
C3-2	61.70	15.00	8.28	4.36	2.73	2.20	5.07	1.29	0.19	0.20	435	27.70	75.2	21.9	48.6	84.3	147
C3-3	61.50	14.90	8.27	4.34	2.72	2.19	5.08	1.28	0.19	0.20	417	27.30	72.0	21.2	50.1	83.0	145
C12-1	60.00	15.10	8.70	4.96	3.18	2.02	4.87	1.44	0.20	0.22	401	33.00	88.2	26.9	59.9	102	146
C12-2	60.00	15.10	8.68	4.95	3.18	2.04	4.87	1.42	0.20	0.21	400	32.80	90.5	27.3	60.8	102	146
C12-3	59.90	15.00	8.78	4.93	3.23	2.02	4.83	1.44	0.20	0.22	400	33.10	96.0	26.9	62.1	102	146

Chemical composition of top ash flow deposit from the Eyjafjöll 2010 eruption, 500 m north of terminal end of Gigajökull. Major oxides in w/w %, trace elements in ppm. Analysis: Vattenfall, DK (ISO-9001 certified ash characterisation laboratory). The two triple analytical replications described in the text are shown (C3-1, C3-2, C3-3 and C12-1, C12-2, C12-3..

Table 3. Analysis of variance (ANOVA) of field experiment data from Table 2. s^2 = variance, *F* = F-test, *Pr* = cumulative probability of the F-distribution at the value of *F*, *sr* = relative standard deviation (with respect to the mean). Explanation of subscripts: *ref* = commercial samples, *samp* = field samples, *a* = results from the two triplicate analyses and their pooled variance, *a* = analytical, *samp* = primary sampling. If *Pr(F)* is below 0.95 the sampling variance difference is not significantly different from zero at 95% confidence level.

	SiO_2	Al_2O_3	Fe_2O_3	CaO	MgO	K_2O	Na_2O	TiO_2	MnO	P_2O	Ba	Co	Cr	Cu	Ni	V	Zn
s^2_{ref}	1.10	0.049	0.23	0.64	0.22	0.015	0.026	0.055	0.00023	0.0021	545.0	9.10	572.7	10.0	541.4	254.3	499.5
s^2_{field}	0.264	0.0089	0.0196	0.0258	0.0453	0.0036	0.0075	0.0024	0.00001	0.00007	54.01	3.17	71.0	2.57	40.1	36.5	9.34
s^2_{a1}	0.0233	0.0033	0.00023	0.00013	0.0001	0.00003	0.00043	0.00003	0.0000	0.0000	108.0	0.16	4.48	0.26	5.00	0.56	4.00
s^2_{a2}	0.0033	0.0033	0.0028	0.00023	0.00083	0.00013	0.00053	0.00013	0.0000	0.00003	0.33	0.023	16.1	0.053	1.22	0	0
s^2_a	0.0133	0.0033	0.0015	0.00018	0.00047	0.00008	0.00048	0.00008	0.0000	0.00002	54.2	0.092	10.3	0.158	3.11	0.28	2.00
F	23.1	3.49	13.14	191.7	92.8	48.6	14.5	34.3	0	0	1.21	40.1	6.80	24.7	12.1	169.5	4.23
Pr(F)	0.994	0.821	0.987	1.000	1.0	0.99	0.991	0.997	0	0	0.455	0.998	0.961	0.992	0.987	1.000	0.924
s^2_{samp}	0.251	0.0056	0.018	0.026	0.045	0.0035	0.0070	0.0024	0.00001	0.00007	0	3.08	60.77	2.41	37.00	36.24	7.34
sr_a/%	0.19	0.38	0.46	0.29	0.73	0.43	0.44	0.67	0.00	1.96	1.79	1.00	3.90	1.63	3.16	0.57	0.97
sr_{samp}/%	0.82	0.50	1.58	3.49	7.03	2.79	1.70	3.64	1.59	3.58	0	5.89	9.45	6.72	10.7	6.75	1.87

fully aware, however, that this seeming parallel between the analytical process and the sampling-and-analysis process are, in fact, not comparable at all. This situation has been described in detail in Reference 9. The central issue pertains to what are the consequences of indiscriminate use of different samplers, with varying relevant competences. In an industrial context this is often permitted for "compelling" economic, efficiency or logistical reasons. In such a situation there is an impression that sampling is but simple "bulk materials handling", with only little focus on competence and education. Indeed, industrial primary sampling has often been subject to company *outsourcing* for extremely narrow-sighted budgetary reasons alone, while such scientific focus that may exist is typically more concerned with the type or brand of the equipment, the number of "replicates", or the *analytical method* s.s. All of which are indeed important factors, but which dwindle almost to insignificance when compared to the total effect of the many neglected sampling errors that go unnoticed without proper TOS insight.

Data analysis

The ten individual ash analytical results will be compared to the two available types of reference material results for the same target, Tables 1 and 2, and also graphically by use of multivariate data analytical *projections*, Figures 6 and 7. For the latter, PCA is employed. Suffice here to note that a principal component graphical plot allows easy visual comparison of the correlation between any number of variables; in the present case for a total of 17 variables (10 major element oxides and 7 trace elements), for any number of samples. PCA *projection* depicts the degree of *similarity* between samples by means of their Euclidian distances in so-called *score plots*, which is a reflection of their simultaneous compositional relationships to one-another (in PCA plots *closely* positioned samples have *closely* similar compositions for most, or all, of the variables involved). The sample disposition is "explained" by a complementary projection visualisation of the variable correlation relationships, in the corresponding *loading plots*. See Reference 4 for a full introduction to projection-based multivariate data analysis.

There is a critical aspect of the official data published in Table 1 that merits further discussion, and which well illustrates the themes taken up here. This concerns the use of "averages of duplicate analysis". What was *de facto* duplicated here: the analysis alone? Did this include sample preparation? Did it include sub-sampling? Or was it the primary field sampling? The latter options are highly unlikely in context, but there is no relevant information to be had. In a narrow analytical context, such issues are often not accorded further explanation, but this issue *must* be fully described, lest the users of the analytical results are unable to understand *from where* did the *averaged* analytical differences originate? What caused the analytical variability? It should be made perfectly clear that a comprehensive understanding of such ambivalent replication information is critical with respect to assessing the total sampling-and-analysis uncertainty, i.e. from *which level* in the full sampling-sub-sampling-sample-preparation-analysis pathway did the "duplication" (or "replication") take place? This issue is fully outlined, e.g., in References 4 and 8.

While the date of eruption, not necessarily the same date as that of sampling, plays an important role in understanding the overall evolution of the volcanic system over time and its intermittent eruptions, we will here leave this aspect to the geologists proper. Rather, focus is on the nature of the sampling methods (grab vs composite), which can be of equal influence on the total uncertainty budget, if not more so. By virtue of the experimental design, every effort was made here to sample only the singular, latest, top ash layer, so as to minimise such possible lava compositional differences. It is, therefore, the considered opinion that whatever differences in Table 2 owe their manifestations to the combination of ash fall heterogeneity × sampling representativity. It will not be possible to separate the influence from these two factors from another.[4,8] This is an important general condition for field sampling—identical for many other lot and material types.

From the analytical results in Table 2 it is possible to estimate the analytical variance and the primary sampling variance respectively. The latter reflecting the difference between sampling teams for all elements (except Al, Mn, P, Ba and Zn). The result of an Analysis of Variance (ANOVA) decomposition analysis of the data in Table 2 is given in Table 3.

The most obvious information in Table 3 is that the analytical precision is very good (all relative analytical standard deviations fall between 0.2 % and 4 %), attesting to excellent laboratory performance. By the same token, all ash-fall samples appear to be of a somewhat stable composition (relative sampling standard deviations range only 0.5–9.5 %). It is very illuminating to note that the primary sampling is responsible for the largest variance component in the experiment, because *average* sampling standard deviations compared to the analytical standard deviations in Table 3 amounts to *ca* 85 % across all chemical variables. Sampling process effects very nearly always dominate.

Thus, even if the ash fall in the designated experimental area after analysis turns out to be a much more uniform material than what was originally expected, there are clear, significant inter-sampler (inter-personal) variations. The largest contrasts relate to the elements present at the lowest concentration levels, e.g. MgO (present in an unusual low concentration for a major element) and the trace elements, fully as expected from geology and geochemistry.

Sampling case study: embedded replications

Several experimental design twists can be found behind the data in Table 2. There are three replicated analytical results pertaining to one primary field sample (C3), intending to capture the magnitude of the laboratory sub-sampling effects (samples Nos 3–5 in Table 2). Also, samples Nos 13–15 represent a similar three-fold analytical replication of the second commercially acquired "tourist sample". Alas, the tourist samples come without *any* reported sampling details (no sampling date, no locality information, no sampling method specification). However, it is more than likely that they were acquired by *grab sampling*.

Both these triplications were made intending that the necessary aliquot sub-sampling operations to be included and evaluated together with the analysis (these sample triplications were not known to the laboratory). Thus, these triplicates were prepared for analysis by identical procedures to the ten field samples. These "triplicate analytical results" sets are compared in full in the multivariate data analysis below. The two "tourist samples" (C12 and C13/14/15, respectively) are also of interest because there would appear to be no reason to distrust their authenticity in view of their price in the Iceland souvenir kiosks! Tourists must, of course, be able to trust that they do indeed represent *bona fide* ash samples originating from the Eyafjallajökull explosive eruptive phase.

These commercially acquired samples therefore mimic a very often occurring general problem: "Samples" are required to be analysed, but little, or no, information exists as to their provenance and the specific sampling and sub-sampling employed. How often is an analytical laboratory not faced with this problem, and how often does the laboratory not simply *assume*, or is forced to assume, that the "samples" have been acquired by a representative sampling process. Or the analytical laboratory may opt to take the obvious, easiest way out: "Ours is not the responsibility to question the representativeness of samples received—our job to *analyse* the samples (only)".

Below is a multivariate analysis of all the available chemical data relating to the sampling field experiment and these relevant comparisons. It is not possible to give a full presentation of all the necessary elements of multivariate data analysis, but the specific visualisation issues lend themselves to easy understanding. A complete introduction to multivariate data analysis, specifically PCA, can be found in Reference 5.

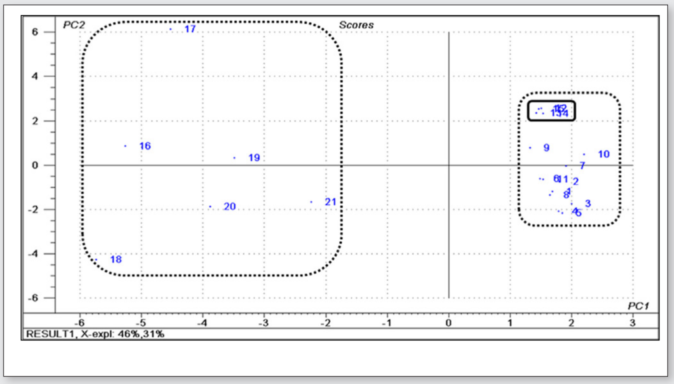

Figure 6A. PCA multivariate projection analysis (17 variables) for 11 field sample analyses (nos 1–11), four commercial analyses (nos 12–15) and six published analyses (nos 16–21).The major difference concerns the field/commercial samples (right) vs published samples (left). The first two principal components model (46 % + 31 %) = 79 % of the total data set variance.

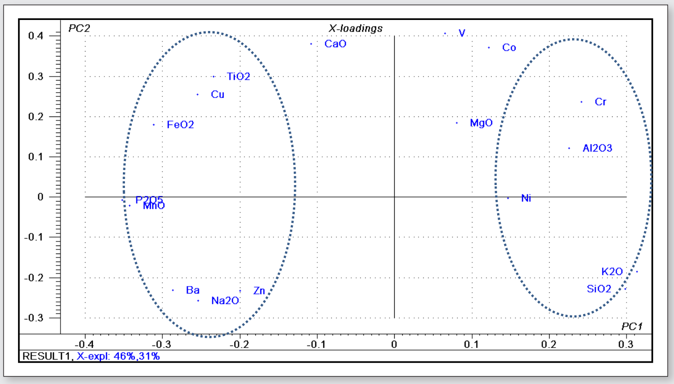

Figure 6B. Variable correlation pertaining to Figure 6A. From standard PCA interpretation, it is observed that the published analytical results are relatively enriched in the oxides of the elements in the left-hand side of this loading plot, while the field experiment samples are relatively enriched in the elements in the right-hand side. The first two principal components model (46 % + 31 %) = 79 % of the total data set variance.

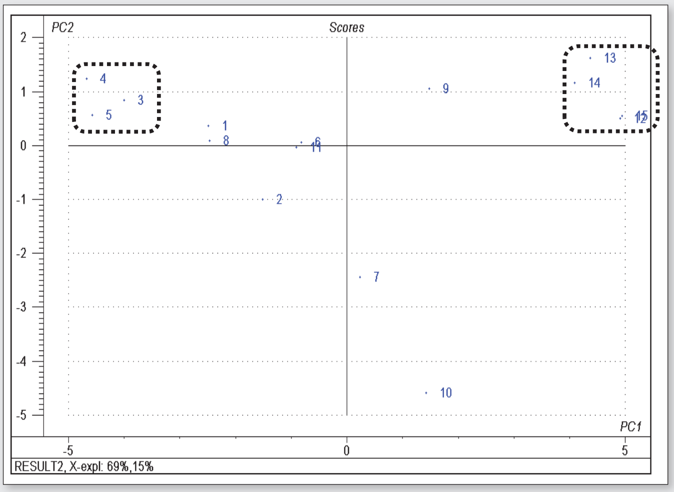

Figure 7A. Focused relationship between field samples and their commercial equivalents only. Commercial sample no. 12 was analysed once, while analyses 13–15 represent triplicate sub-sampling and analysis in the laboratory of the second commercial sample (score box right). Field sample 3 was similarly sub-sampled and analysed in triplicate (analyses 3, 4, 5) in the laboratory (square box right). These embedded replications illustrate the maximum laboratory aliquot handling-and-analysis error, with which to assess the sampling variability between all ten sampling group results. The first two principal components model (69 % + 15 %) = 84 % of the total data set variance.

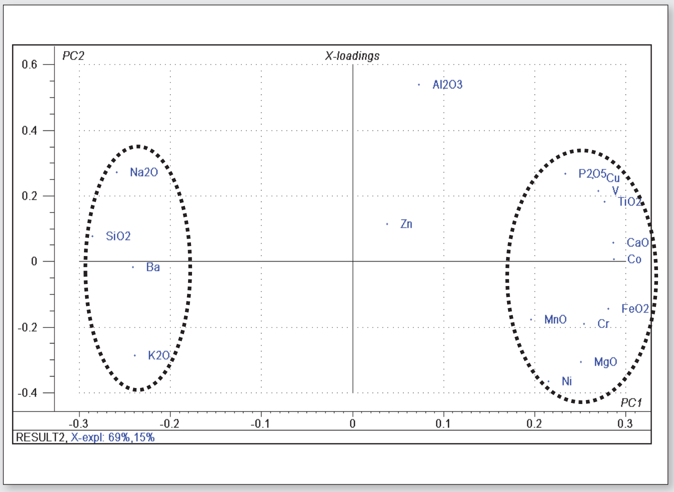

Figure 7B. Correlation relationships between variables for field and commercial samples in Figure 7A. The “silicic” (SiO_2, K_2O, Na_2O, Ba) vs “mafic” composition (MgO, CaO, FeO_2, Ni, Cr ...) differentiation is a well-known geochemical differentiation feature of basaltic volcanic lavas. The first two principal components model (69 % + 15 %) = 84 % of the total data set variance.

Technical note: for the data analytical cognoscenti

All variables in the principal component analysis (PCA) have been *auto-scaled*, i.e. *centred* (w.r.t. the average of all variables) and *normalised*, i.e. divided by 1/std. This facility allows data differences to be augmented with the utmost clarity in a completely balanced fashion. Auto-scaling is a mandated first data transformation in general PCA. For further details see, for example, chapter 5 in Reference 5.

Sampling case study: multivariate data analysis and visualisation

Figure 6 is a multivariate PCA *projection* rendition of the data in Table 1 and 2 (*score plot*), showing the mutual relationships of all field samples at a glance. The commercial samples contrast markedly with the six published ash compositions. These two main groups are delineated; the commercial samples are labelled and distinguished within the field experiment group (right).

Based on this visualisation it can readily be observed that these earliest published analytical results are markedly different from the field and commercial results, as witnessed by the left vs right sample clusters respectively. This represents the geologically known evolution of the general lava composition over the entire eruption duration.

If one wishes to study further why this is so, in the “language” expressed by the variable correlations, the accompanying *loading plot* (Figure 6B) outlines those variables for which these two sample groups display relatively high vs low concentration levels. Thus, the published results are relatively high in the concentration of elements FeO_2, MnO, P_2O_5, Ba, Cu, Na_2O, Zn and TiO_2, while the field and commercial samples display relatively low concentration levels in these elements. Some, but not all, of these relationships can be easily reconciled with standard geochemical basaltic magma differentiation understandings, but this issue need not be pursued further here.

For the present purpose we proceed to study in more detail the relationship between the field experiment samples and the closely related commercial samples only. These focused relationships are displayed in Figure 7A and B.

In Figure 7A (*score plot*) all field samples and their commercially counterparts can be readily compared in the most discriminating fashion in comparison with the absolute analytical results shown in Table 2. It is notable that field sampling group 3 sampled material which is maximally different from the commercial compositions (this difference can be put on a quantitative basis by reference to Table 1), with the preponderance of results from most of the other field sampling groups *lining up* along principal component No. 1, PC_1. Ordering this polarity feature with respect to decreasing similarity with group 3, ranks sampling groups Nos. 3, 1, 8, 2, 6, 11, 9.

The most notable among the ten field sampler groups would appear to be groups 7 and 10, which set themselves off in a different fashion, such that this feature is modelled by the second principal component, PC_2. When interpreting complementary score, and loading plot relationships, the relative proportions of the total data variance (in Table 2) must always be kept in mind, in the present case the PC_1 variability account for more than four times the variance modelled along PC_2 (69 % vs 15 %, respectively). Thus, the deviations reflecting groups 7 and 10 must be seen, and assessed, in this moderating light.

Figure 7A also delineates the variability due to the two laboratory triplicate aliquots replications, intended to reveal the sum-total of the specific aliquot extraction-plus-analysis error effects. Figure 7A reveals that the variability of these laboratory uncertainties (square boxes) are only responsible for a minor part of the total between-sample variability along both PC_1 and PC_2 axes, which collectively accounts for 84 % of the total data variance. Variations larger than these box dimensions must therefore be ascribed to differences between the ten sampler groups, given the information from Tables 1 and 2 that the local ash fall is of relatively stable composition.

It is highly relevant also to take a look at the variable relationships corresponding to Figure 7A (*score plot*), which are given in Figure 7B (*loading plot*). This data set gives rise to a practically 100 % understandable variable correlation signature for all 17 variables, fully consistent with conventional volcanic and geochemical knowledge: The "silicic" composition (SiO_2, K_2O, Na_2O, Ba) vs the anti-correlated "mafic" composition (MgO, CaO, FeO_2, Ni, Cr) polarity is a well-known geochemical feature of basaltic volcanic lavas, which lends credibility to the relevance and quality of the sampling undertaken.

Figure 8. A sampling group with "2 × 60 years", in field and birthday party mode.

Discussion and conclusions

As to the *confounding* factor described above, it is not possible to distinguish with complete resolution between the effects from material heterogeneity and the somewhat different composite sampling procedures used (from 3 to 10 increments for the specific composite samples). Indeed, there is also the possibility that one or more of the sampler groups accidentally tapped into the *sub-drill* material, which will have affected the composite sample compositions to a varying, but significant degree. Indeed, this *could* have happened for each individual increment used. Could such, for example, be the reason behind the most deviating sampler groups 7 and 10?

The value of incorporating relevant elements of the discipline of Design of Experiments (DoE) was illustrated in the results interpretation above, e.g. the value of a small embedded experimental design of triplicate replications of the final aliquot sampling step. DoE is most often applied in situations in which the experimental factors can be fully controlled, but it can also be creatively carried over to many other scenarios. Thus, the present simple aliquot replications showed up clearly in the multivariate projection plots, allowing to take in a measure of discrimination between the specific analytical uncertainty effects vs the complete between-sampler group variability.

The reason this article has traced what may appear as very detailed geological issues in such detail is that these are actually, or mimic well, very general features that attend *all* primary sampling of significantly heterogeneous materials in practice. Many lots display a similar degree of complex origin (temporal, spatial), also, lots defined as targets for environmental studies, natural processes, biological materials within agriculture, wine making etc.[10]

By multivariate PCA projection it was possible to obtain a complete overview of all essential similarities and differences between samples and in the present specific case, also between all field sampling groups.[b] These groups are of course anonymous, and shall remain so, only identified by numbers. It matters not what is the identity of the sampler groups shown in Figure 2 and the corresponding results shown in Figure 7. What matters is the degree of sampling variability shown by a group of interested, willing and eager samplers, but with significantly varying TOS competence and experience difference after all other factors have been optimally controlled. The realism of the experiment is poignant and highly relevant. Field sampling of lots of complex origin and significant heterogeneity is no walk in the park; TOS is critically needed for all *sampling processes* that aspired to representativity—regardless of the specific nature of the target lot.

What also matters is the *possibility* of augmented interpretations by taking into account all variables simultaneously—no less than 84 % of the total variability for 17 variables is captured by the graphical

---

[b] In this context, it is perhaps of relevance that field sampling groups 2, 3 and 9 are the groups supposed to know their geology and sampling business well.

rendition based on only the two first principal components. What is evident from this data analysis is that there is *prima facie* close general similarity between all the field samples and the commercially obtained samples, Figure 6, but also that it is still possible to trace out and interpret highly detailed relationships *between* alternative sampling groups, a few of which would appear to perhaps have gone somewhat astray (or more likely, simply "sampled a bit too deep") based on what in "absolute" concentration units (percentage points) constitute only very small differences, cf. Tables 1 and 2. Such is the power of relevant multivariate data analysis (*chemometrics*), which in some aspects rather is a hidden elephant in the room. While many sampling tasks can be optimised by identifying the singular variable (compound, analyte) for which the lot heterogeneity is the largest, there also exist many other cases in which this is not known in beforehand, and where a simultaneous overview of the full variable correlation make-up can yield interesting pathways to new understandings and scientific/technological problem formulation.

Lessons learned

1) The Eyjafjellajökull field experiment illustrates sampling of a non-industrial heterogeneous lot with a complex origin, viz. the geological formation of the local 2010 volcanic eruption ash fall. This lot displays many potentially confounding characteristics and features. The specific field activity shows how a dedicated experiment can be designed for those sampling factors and sampling conditions that can controlled individually—laying bare only the ultimate, unavoidable confounding between degree of lot heterogeneity × sampling method.
2) There is power in adding the discipline of DoE to the professional sampler's arsenal, allowing to *design* the entire sampling campaign and/or place *embedded* experiments within any sampling superstructure, see, for example, chapter 11 in Reference 5.
3) The versatility of the RE approach could be illustrated with particular clarity. Regardless of how many sampling factors, conditions and stages involved before analysis, a RE "from the top", i.e. replicating the primary sampling will *per force* pick up effects from all active sampling errors (from all sampling stages), which is exactly what is the objective. The RE is described, for example, in chapter 9 in Reference 5 and in Reference 8.
4) Multivariate data analysis (*chemometrics*) is a further empowerment for samplers, allowing the overview, at a glance, of the relative sampling variabilities, e.g., from aliquot-extraction-plus-analysis vs the total sampling variance, Figures 6 and 7. A multivariate data analysis overview will also allow samplers to isolate the candidate analyte with the largest heterogeneity manifestation—and will depict which other analytes are correlated/anti-correlated/not correlated with, allowing a first foray into a multivariate work modus for TOS.
5) Multivariate data analysis allows extracting inherent correlation features between samples and analytes with the outmost clarity and precision, even in the case of great similarity between analytical data due to the convenient data analytical "auto-scaling" facility.

Acknowledgements

Vattendall DK and Vattenfall Europe are thanked for the analytical results of all samples included in this experiment. Guesthouse Red Eric, Reykjavik, is thanked profusely for boarding in Reykjavik and for extraordinary mobile catering during the entire week on Iceland—much thanks to Edda and Runar. Listasafn Sigurdur Olafsson provided a cultural backdrop *extraordinaire* for these activities, much thanks to Birgitta and Hlif. Anne was the critical support and partner in all planning and execution of the complete Iceland birthday trip, much love and many hugs to her. The institution GEUS is thanked for providing a solid foundation for the first author's sampling achievements and results in the very active period 2010–2015, experiments, studies, publications and for a most efficient infrastructure—*there was even the occasional fellow geologist who became interested in sampling, though sadly not many*.

Finally this publication is dedicated to life-long friendships, many of which are featured above; a very special recognition to soul-mate, equally "young", Guttorm Isaksson of Tromsø fame.

References

1. M.T. Gudmundsson, R. Pedersen, K. Vogfjörd, B. Thorbjarnadóttir, S. Jakobsdóttir and M.J. Roberts, "Eruptions of Eyjafjallajökull Volcano, Iceland", *EOS* **91(21),** (2010). https://doi.org/10.1029/2010EO210002
2. F. Sigmundsson, S. Hreinsdóttir, A. Hooper, Th. Árnadóttir, R. Pedersen, M.J. Roberts, N. Óskarsson, A. Auriac, J. Decriem, P. Einarsson, H. Geirsson, M. Hensch, B.G. Ófeigsson, E. Sturkell, H. Sveinbjörnsson and K.L. Feigl, "Intrusion triggering of the 2010 Eyjafjallajökull explosive eruption", *Nature* **468,** 426–430 (2010). doi: https://doi.org/10.1038/nature09558
3. R. Pedersen, *Eyjafjellajökjull: Vulkanen der Lammede Europa*. Gyldendal (2014). ISBN: 978-87-02-10062-4
4. K.H. Esbensen, "Pierre Gy (1924–2015): the key concept of sampling errors", *Spectrosc. Europe* **30(4),** 25–28 (2018). https://www.spectroscopyeurope.com/sampling/pierre-gy-1924%E2%80%932015-key-concept-sampling-errors
5. K.H. Esbensen and B. Swarbrick, *Multivariate Date Analysis—An introduction to Multivariate Data Analysis, Process Analytical Technology and Quality by Design*, 6th Edn. CAMO Software AS (2018). ISBN 978-82-691104-0-1
6. *DS 3077. Representative Sampling—Horizontal Standard*. Danish Standards (2013). www.ds.dk
7. *Get Your Sampling Right: An entry-level introduction to representative sampling*. A collection of the Sampling Columns published in *Spectroscopy Europe* (2018). https://www.spectroscopyeurope.com/get-your-sampling-right
8. K.H. Esbensen and C. Wagner, "Sampling quality assessment: the replication experiment", *Spectrosc. Europe* **28(1),** 20–25 (2016). https://www.spectroscopyeurope.com/sampling/sampling-quality-assessment-replication-experiment
9. K.H. Esbensen and C. Wagner, "Theory of Sampling (TOS) vs Measurement Uncertainty (MU) – a call for integration", *Trends Anal. Chem.* **57,** 93–106 (2014). https://doi.org/10.1016/j.trac.2014.02.007
10. K.H. Esbensen, "Theory of Sampling – what's next?", in *Proceedings of the 8th International Conference on Sampling and Blending*. AusIMM, pp. 29–38. (2017). ISBN: 978-1-925100-56-3