Journal of Structural Geology 32 (2010) 818–831

Contents lists available at ScienceDirect

Journal of Structural Geology

journal homepage: www.elsevier.com/locate/jsg

Statistical analysis of an experimental compressional sand wedge

Nadaya Cubas [a,1], Bertrand Maillot [b,*], Christophe Barnes [b]

[a] École Normale Supérieure, CNRS, Laboratoire de Géologie, F-75005 Paris, France
[b] Université de Cergy-Pontoise, Département Géosciences et Environnement, F-95000 Cergy-Pontoise, France

ARTICLE INFO

Article history:
Received 9 October 2009
Received in revised form
19 May 2010
Accepted 21 May 2010
Available online 9 June 2010

Keywords:
Analogue modeling
Statistical modeling
Accretionary wedge
Experimental error bar

ABSTRACT

The quasi-static deformation of dry sand is widely used as an analogue to the brittle deformation of the upper crust. The quantitative comparison of analogue to natural tectonics, or to mechanical predictions, requires identifying sources of biases and estimating the intrinsic variability of the experimental results. We develop experimental and statistical methods that fulfill these requirements. We consider an initially perfect wedge resting on a flat layer, made of a uniform dry sand in a rectangular glass box. The box is shortened lengthwise by translating one of its end walls towards the other. The lateral walls can remain fixed, or be translated with the moving end wall. Upon shortening, the wedge is thrusted above the flat layer forming classical fore- and backthrusts, as essentially plane-strain, structures. Lifetimes, locations, and dips of all thrusts constitute seven quantifiable output parameters (called observables), in addition to the shortening forces monitored at both end walls during shortening. Up to seventy measurements of each observable were performed in seven final-state cross-sections of ten experiments. A three-step statistical analysis allows us to prove that, first, the observables vary independently, justifying their modeling with independent distributions. Second, the ergodic hypothesis holds, meaning that along strike variations can be used to infer the intrinsic experimental variability. Measurements can thus be repeated on successive cross-sections in each experiment. Third, our data set is free from bias due to friction on the lateral walls, or due to the finite length of the box. We then construct statistical models of each observable using either Gauss or Laplace distributions. For example, forethrusts dip at $38^{\circ} \pm 3.2^{\circ}$, and backthrusts, at $41^{\circ} \pm 3.3^{\circ}$. We finally show how to apply these statistical models to experiments using a different initial geometry. The statistical methods presented here are applicable to experiments with different setups, materials and observables, although the ergodic hypothesis is relevant only to plane-strain experiments.

1. Introduction

In the study of real tectonic structures, sand-box techniques are used to produce analog deformation histories from undisturbed sedimentary strata to final states that resemble actual structural geometries. Cadell (1888) provides one of the earliest examples of a mountain building due to horizontal shortening above a décollement, while the cross-section of the Northern Alps in granular materials made by Bonnet et al. (2007) is one of the latest and most impressive examples. Koyi (1997) provided a useful historical review of these experimental techniques.

Comparisons of the results of sand-box models to real structures are often qualitative rather than quantitative, owing to structural complexity and to limitations concerning the quantitative description of the results of sand-box models. Attempts at developing quantitative interpretations started with King Hubert (1951) who showed that the quasi-static deformation of dry sand is consistent with the Coulomb failure criterion, opening the way for sand-box modeling of tectonic structures. Later, sand-box modeling of a wedge was used to validate theoretical predictions of critical wedge theory about the shape and evolution of an accretionary wedge (Davis et al., 1983). Further, the nature of accretion by underthrusting was successfully compared to work estimations using a boundary-element method (Del Castello and Cooke, 2007), and the dip of backthrusts in the wedge predicted from an analytical force balance with the principle of minimum dissipation was validated (Maillot and Koyi, 2006). During this time, both discrete element models (Saltzer, 1992; Seyferth and Henk, 2006; Egholm et al., 2007; Hardy et al., 2009) and finite-element models (Ellis et al., 2004; Crook et al., 2006) were used to conduct comparisons with sand-box experiments. Discrepancies between various numerical techniques and implementations have been demonstrated by benchmark tests (Buiter et al., 2006).

* Corresponding author.
E-mail address: bertrand.maillot@u-cergy.fr (B. Maillot).
[1] now at Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA.

0191-8141/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsg.2010.05.010

In general, to have predictable quantitative experimental results, reproducibility must be prevalent. To our knowledge, the only attempt at quantitatively considering reproducibility is the benchmark tests by Schreurs et al. (2006). Ten different laboratories performed two plane-strain experiments, respectively in contraction and extension. Large discrepancies were observed in the positions, dips, and numbers of thrusts. These variations were attributed to differences in shear along the lateral walls of the boxes, in granular materials and in their handling methods. The problem of the relative movement of the lateral walls with respect to the moving end of the box was well described during a discussion between Koyi and Cotton (2004) and Costa and Vendeville (2004). Significant experimental artifacts have also been demonstrated when basal friction is very low compared with lateral friction (Vendeville, 2007). Here, we will explicitly demonstrate that our experimental data are free from such biases. Additionally, by using the same sand and the same handling protocol in all experiments, we eliminate the main biases identified in Schreurs et al. (2006) for our quantitative analysis of sand-box model behavior.

The aim of this article is to quantify the experimental variability and to identify sources of experimental biases. We considered four different experimental prototypes, or setups, and we repeated experiments for each setup. All setups consist of a pre-built sieved sand wedge occupying all the box width and approximately a third of its length. The rest of the box is filled with a flat sand layer. Differences in the setups concern only the relative slip sense of the lateral walls, the surface slope of the wedge and the dip of the base of the box (Table 1). We study the growth of the sand wedge by frontal accretion for an overall shortening of the box of about 8% generating one or two forethrusts in a forward sequence. We then construct statistical distributions describing the variations of the observables (i.e., positions, lifetimes, dips of the thrusts, and the compressive forces at both ends of the box). Thus, the experimental results are described in a probabilistic framework, with the ultimate goal of inverting them to retrieve mechanical properties of the sand, as was done in a simpler experimental setup by Maillot et al. (2007) and for a case study of Nankai wedge (Cubas et al., 2008). These setups are very close to that proposed by Schreurs et al. (2006). In this respect the present study complements theirs, by repeating the experiments using the same protocol, sand, and box, instead of comparing the results of different laboratories, and by introducing a statistical treatment of the experimental results.

2. Experiments

The experiments consist of shortening by a fixed amount of an initially perfect sand wedge resting on a flat sand layer (Fig. 1a). To accommodate the shortening, the sand is lifted in the flat layer at the front of the wedge, following a classical fore–backthrust geometry. Overall shortening magnitude was chosen to be sufficient to generate two fore-back thrust pairs (Fig. 1b). However, the intrinsic variability of the experiments yields the result that two experiments out of ten produce only one thrust pair (Fig. 1c).

2.1. The sand-box

The experimental box is rectangular, with inside dimensions of 280 mm (width), 370 mm (length) and 90 mm (height). It can be tilted lengthwise producing a dip of the basal plate in the direction of shortening, which is typical of accretionary wedges. All the box is built with 10 mm-thick glass with a precision of $\pm$0.05 mm, which is an error smaller than our smallest grain sizes. In all experiments, shortening of the sand is imposed by translating the back wall (i.e., the wall on the wedge side) over 30 mm at a rate of 0.52 mm/s with an electric motor via a screw.

This sand-box allows us to infer the effect of the horizontal shear stress due to sand friction on the lateral walls of the box by using two box configurations. In configuration A, the lateral walls are translated with the back wall, while in configuration B, the back wall moves alone (Fig. 2a,b). Any plane-strain sand-box experiment fits into one of these two categories. Sometimes in the literature, configuration A is referred to as a “pull” set up that is typical of boxes using a conveyor belt, and configuration B as a “push” set up. Note that in configuration B, the sand is overall lifted up by the movement of the back wall, while in A, the back wall remains at a constant altitude. Here, the box is horizontal ($\beta = 0$), so the only difference between the two configurations is the lateral friction. Basal and side glass walls were treated with a carbon based product (“RainX”) that reduces friction with sand. During the experimental work, we switched the lubricant for a new “RainX” release from the maker (Shell International). The change in lubricants did have a slight effect on friction and cohesion, which is quantified in Section 3.3. We measured the friction of the sand for the glass treated with the new “RainX”, and obtained a friction angle of 7.5°–10° (coefficients of 0.13–0.18) and a negligible cohesion of about 10 Pa. For comparison, the untreated glass yielded 12°–18° (0.21–0.32), and the same negligible cohesion.

A second feature of the box is that strain gauges are behind the end walls and measure the force that the walls sustain during shortening (Fig. 1b,c). A piece of cloth was placed between the end and lateral walls to prevent blockage due to sand leaks and thus improve the precision of the force measurements. Vertical movement is free but the weight of the glass wall resting on the bottom plate is sufficient to prevent its uplift, although a very thin film of sand (a fraction of a mm) still leaks below the moving back wall. Before filling the box with sand, a shortening of 30 mm is applied so that all parts of the box adjust their respective positions. In particular, the strain gauges adjust their tension so as to equilibrate friction of the back wall against the rest of the box. We then set the zero signal of the gauges in this very configuration and completed filling of the box with sand. Thus, during further shortening, the gauges measure only the additional force needed to shorten the sand, and not that needed to move the walls. It was difficult however to ensure a constant tension of the gauges during box filling, and this part of the experimental protocol may be at the source of a bias for the force measurements, which will be discussed in Section 4.

Table 1
Experimental prototypes.

Prototypes	L (mm)	H (mm)	l (mm)	h (mm)	α (deg)	β (deg)	S mm	Wall config.	Number of experiments
1A	370	26	162	9	6°	0°	30	A	5
1B	370	26	162	9	6°	0°	30	B	5
2A	370	26	96	9	7°	3°	30	A	2
2B	370	26	96	9	7°	3°	30	B	2
3	407	40	176	15	8°	0°	43.5	A	1

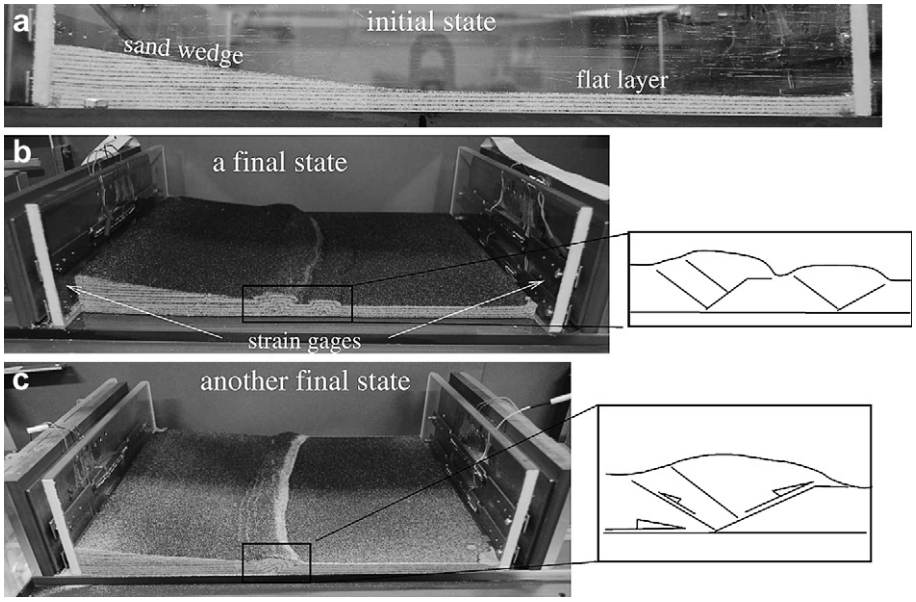

Fig. 1. Photographs of initial (a) and typical final (b,c) states of the experiments. Note the opposite curvature of thrusts on the topography in (b), where the lateral walls remained fixed (Fig. 2b), and in (c), where they followed the translation (Fig. 2a). The strain gauges placed behind the walls (b) allow us to measure the horizontal compressive force at both ends of the box during the shortening.

2.2. The sand body

A layer of dry sand of thickness $h = 9$ mm covers the inside of the box and is overlain by a wedge of the same sand with surface slope α and maximum thickness $H = 17$ mm at the back wall (Fig. 3a). The length l of the wedge depends on its slope α and basal dip β as $l = 17/\tan(\alpha + \beta)$ mm (Table 1, Fig. 3a). A displacement $S = 30$ mm was applied in all experiments.

We only used a single batch of sand for all the experiments: Fontainebleau aeolian quartz sand (99.7% of quartz) of median grain size 250 μm with 95.5% in mass of grain sizes comprised between 150 and 425 μm. The sand was previously found to have a peak friction of 33° weakening to 30° over a slip of 3 mm for a newly formed slip surface, and 1 mm for reactivation of an existing slip surface (Klinkmuller et al., 2008). The sand pack was produced with a sand distributor that was adapted from Wygal (1963) to improve the planarity of the pack (Fig. 4). Sand packing is dense, reproducible, and homogeneous throughout the box with a value of 1710 ± 6 kg/m^3. During our experiments, the temperature in the laboratory remained between 17.5° and 22.5°, and the relative humidity, between 60% and 78%.

2.3. Experimental prototypes

Four different experimental setups were used and we refer to them as the prototypes. We conducted a total of fourteen experiments using one of the four prototypes (Table 1).

For prototypes 1A and 1B, the box is horizontal ($\beta = 0$) and the initial slope of the sand wedge is $\alpha = 6°$. For prototypes 2A and 2B, the box is tilted by 3° and the wedge surface slope is 7° so that its taper angle is $\alpha + \beta = 10°$. The letter 'A' or 'B' following the prototype number refers to the configuration of the lateral walls (Fig. 2a,b). Both prototypes of 1 were repeated experimentally five times, and prototypes of 2, twice. Ideally, all experiments for a prototype would yield the same results, but variability did occur, necessitating the statistical analysis that is the focus of this contribution. The analysis focuses on the 10 experiments for the two types of prototype 1. Data from prototypes 2 are only used in Section 4.2 to test whether their variability follows the same statistical behavior determined from prototype 1 experiments, i.e., to test the statistical homogeneity of the distributions obtained from prototypes 1.

2.4. The observables

Generally, after 4 to 5 mm of shortening during the experiments, the sand deforms by the formation of two conjugate reverse faults called 'forethrust 1' and 'backthrust 1' (Fig. 3b). Continued shortening displaces the wedge over the forethrust creating new relief. Backthrusts rooted at the base of the forethrust form regularly to ensure kinematic compatibility during thrusting. A second fore- backthrust system forms beyond the first one during late shortening (Fig. 3c), except in two experiments of prototypes 1B.

Eight "observables", i.e. quantifiable output parameter, were gathered for comparison and characterization of the experimental outcomes. The first observable, named δ (Fig. 3c) is the shortening accumulated at the back wall during displacement on the first forethrust, or the "lifetime of the first forethrust". We measure it by filming with oblique lighting the top surface of sand during shortening and identifying the emergence of the first and second forethrusts. The interpretation of δ as a measure of slip accumulated on the forethrust is supported by the absence of visible diffuse compaction or layer parallel shortening during displacement on the forethrust. It is however possible that compaction accommodated some deformation above the sole thrust ahead of the forethrust ramp (Mulugeta and Koyi, 1987, 1992).

After the end of experimental deformation, the sand was dampened, 15 cross-sections cut at 20 mm intervals across the box width, and the profiles were photographed. On each photograph, we measured (Fig. 3d): the dips γ_1 and γ_2 of the first and second forethrusts, respectively, as well as the dips θ_1 and θ_2 of the associated backthrusts. These dips are visible on the photographs from the geometry of the strain markers in the sand (Fig. 1). Where several backthrusts formed at a forethrust, we measured only the

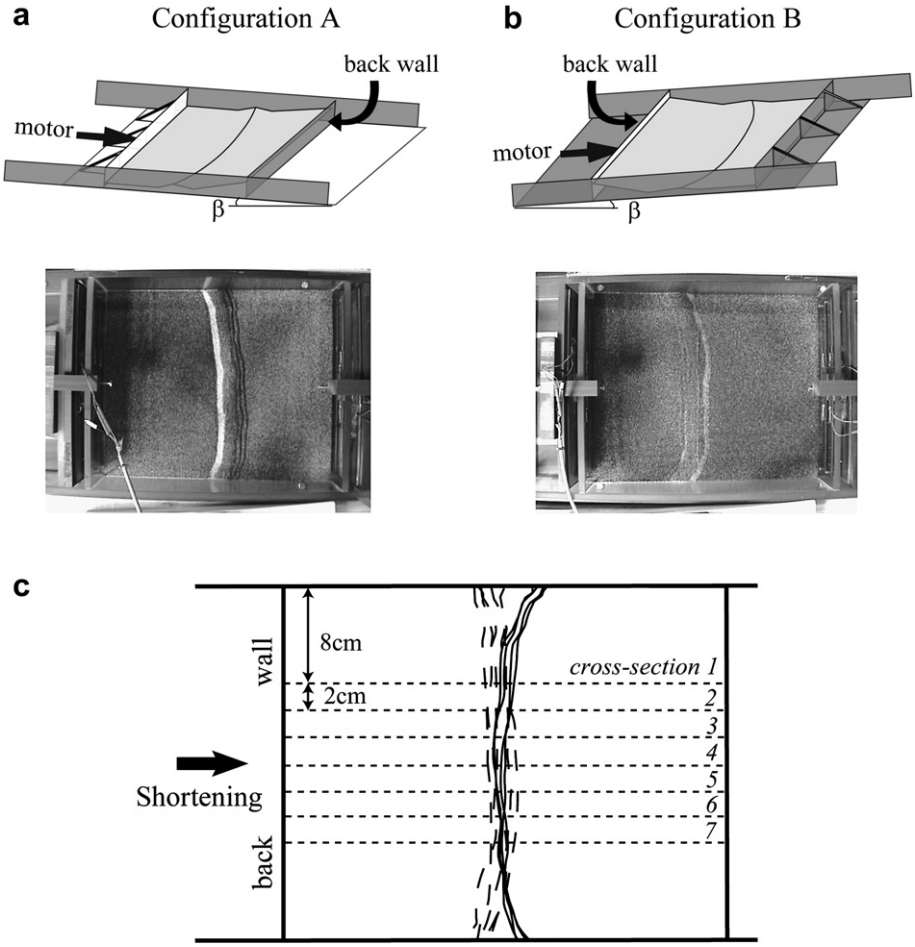

Fig. 2. The two box setups used to identify side-wall effects. In a), the motor pushes together the left hand wall with the basal plate (in white) down dip (angle β) towards the back wall. In b), the basal plate is turned by 180° and the motor now pushes the back wall (in white) alone up dip. All parts shown in grey are fixed with respect to the Earth reference frame. Typical top view photographs at the end of experiments are shown below each configuration. In (c), top view drawings of the emergence of the first forethrusts in the final states of all ten experiments of the prototypes 1A (solid lines) and 1B (dashed lines). To allow comparison, the initial prism is now at the left for both configurations. The thin dashed lines indicate the positions of the seven cross-sections selected to measure the observables.

last one. Second, we measured the lengths d' and r' (Fig. 3), and we define the location of the first forethrust as $d = d' - l$. Location of the second thrust ramp, r, is defined as

$$r = r' - h/\tan(\gamma_1) - h/\tan(\theta_2). \quad (1)$$

Thus, each cross-section is characterized by seven geometrical observables: three lengths (δ, d, r), given in mm throughout the article, and four angles (γ_1, γ_2, θ_1, θ_2), given in degrees.

Repetition of measurements by the same or different observers yielded a typical range of angles of $\pm 1.5^\circ$ and range of lengths of ± 1 mm. These ranges are smaller than the variability of the data and are actually about the size of the symbols used in graphs to illustrate the results (e.g., Fig. 5). This variability is neglected in the statistical analysis.

The eighth and ninth observables are the average forces along the direction of shortening sustained by the front and back walls and measured with strain gages. Only the force on the front wall was used in the analysis of biases (Section 3.3). The force at the back wall is discussed in the section on the statistical modeling of the data.

Finally, the initial slope α of the wedge varies between experiments, and experiences a slight change of about half a degree due to compaction deformation occurring during initial shortening. We measured α at the end of each experiment (noted α_f), and we interpreted it as the actual wedge slope at the onset of thrusting, because it corresponds to the arrest of diffuse compaction.

3. Statistical analysis of the observables

The purpose of the statistical analysis, is to establish whether: (i) the measurements of the observables are independent; (ii) we can interpret each cross-section as an independent experiment; (iii) we can identify biases in the data due to the experimental apparatus and protocol. To determine the answers to these three questions, we used the data about observables from the Prototype 1A and 1B experiments. For these experiments, we used the seven central cross sections to minimize bias due to friction on the lateral walls. Thus, for ten total experiments, seventy total cross sections provided up to seventy measurements of each observable (Table 2, column 1).

3.1. Independence of the observables

Given seven geometric observables, twenty-one graphs of pairs can be constructed (Fig. 5). All graphs show roughly cloud-shaped distributions, indicating that all observables have mono-modal distributions. The values of δ are slightly clustered: they tend to

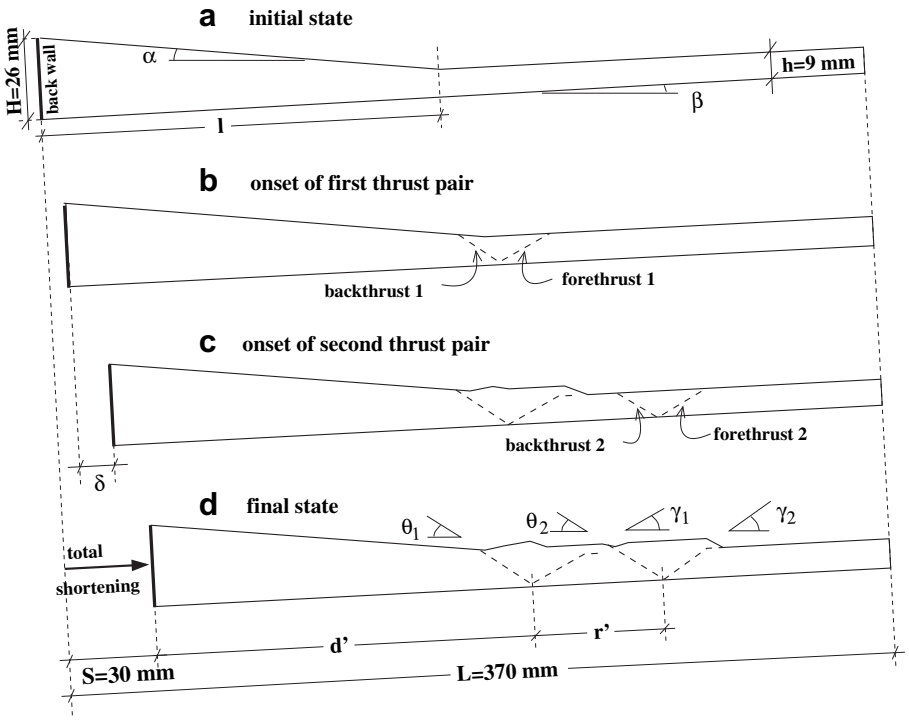

Fig. 3. Illustration of the experimental setup and observables. The initial conditions are defined by L, H, l, h, α, β and the applied shortening is S (Table 1). The observables are the lifetime of the first forethrust (δ, in mm), the location of first and second forethrusts: $d = d' - l$ and r (computed from r' using equation (1)), and the thrust dips θ_1 and θ_2 for the backthrusts, γ_1 and γ_2 for the forethrusts.

gather along horizontal lines in all six graphs where they are plotted. The clusters are populated by data from the cross-sections of the same experiment. Indeed, the films for the top views of the sand during shortening show that the forethrusts emerge at the surface nearly simultaneously across the whole box, thereby yielding almost the same value of δ for all cross-sections of an experiment. Note that the maximum measured value of δ is 25 mm because the first forethrust usually formed after 4 to 5 mm of shortening and the total shortening was 30 mm. In all graphs, the absence of clear elongation of the distributions along any direction is a result of the independence of the observables. Considering that all cross-correlations are below 40%, we conclude that the seven geometric observables are sufficiently independent, particularly because clear dependence is usually concluded for correlations above 75%.

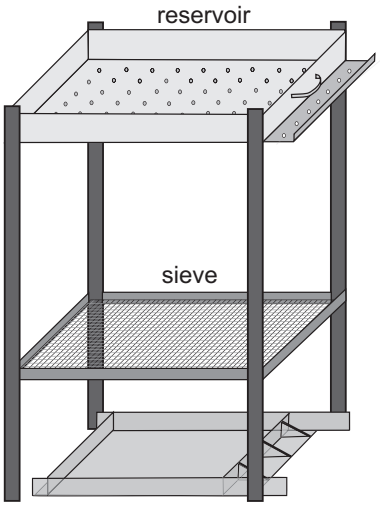

Fig. 4. Schematic illustration of the sand distributor for producing the initial sand pack. Sand flows from the holes at the bottom of the reservoir with dimensions slightly greater than the box. The sand grains are diffused and slowed down by a sieve before falling as a uniform rain into the sand box, and settling at a rate of 0.3 mm/s. Every 2 mm, the sand rain is stopped and a thin layer of colored sand is sieved across the box to act as a marker. The wedge slope is created by 2 mm-thick steps using a cache above the sand box. The only manipulation of the sand consists in filling the reservoir of the sand distributor. No scrapping, vibrating, compressing step, or sharp tap is performed.

3.2. Hypothesis of ergodicity

For reliable statistical testing, experimental measurements must be repeated as many times as possible. If we choose to study a single cross-section in each experiment (e.g., the central one), we will measure only one value of each observable per experiment. Consequently, several tens of experiments would be needed before a statistical analysis was possible. Conversely, if we reduced the number of needed experiments by choosing to study as many cross-sections as practically possible (e.g., cutting sand slices every 2 mm), we risk measuring values that are not independent and hence, not statistically valid. To ascertain the appropriate mix of number of experiments and spacing of sections to create a data population with internal independence for statistical analysis, we need to establish ergodicity, or the statistical equivalence between variation along strike and variations across repeated experiments. We therefore check the two following conditions: (1) two successive cross sections should not be statistically dependent; (2) an observable must vary among cross-sections of any experiment in a manner similar to its variation across the whole data set for all experiments. Because we have at most seventy measurements of any observable, we have recourse to visual tests (Figs. 6 and 7) rather than numerical tests.

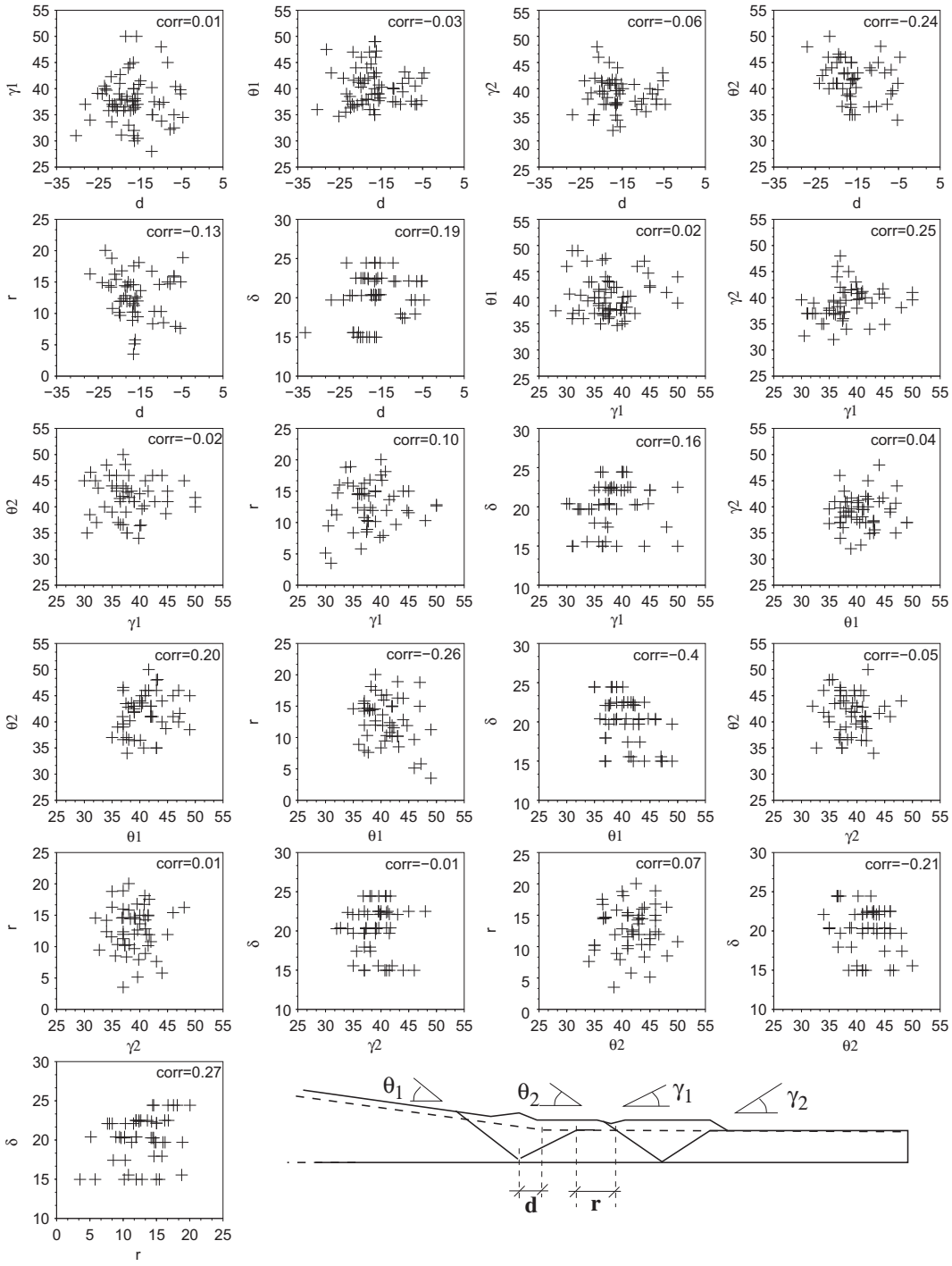

Fig. 5. Graphs of the measurements of each observable versus all others, for the experiments of prototypes 1A and 1B. The observables are illustrated in the bottom right drawing (except δ, Fig. 3c) where the dashed line indicates the initial sand wedge. Correlation between each pair of observables is indicated in the top right corner of each graph.

For statistical independence to apply, values for the geometric observables should display random variations from one section to the next in an experiment (Figs. 2c and 6). Curves in Fig. 6 should thus display irregular saw-tooth shapes. The four graphs for the fault dips (θ_1, θ_2, γ_1, γ_2), and that for r, suggest a satisfactory independence of the cross-sections. The graph for d show less irregular curves, and therefore more dependence between the cross-sections. The graph for δ is markedly different, with almost constant values for all cross-sections in an experiment. In fact, four experiments show a constant value with only four showing a variation in value. A rough estimate of the sufficient distance between sections can be obtained by adding the total widths spanned by all eight curves (8×14 cm), and dividing this by the total number of variations in all curves (5). Thus, substantial variations of δ are expected to occur approximately every 20 cm along strike.

For the second condition, values from a single experiment should span the same range as for all experiments, which seems applicable for the angular observables from visual inspection of the data (Fig. 6). In contrast, d, δ, and to a lesser extent r, show values from a single experiment typically occupying only a portion of the

Table 2
Statistical models for observables of prototype 1 experiments.

	Number of meas.	Number N of indep. meas.[a]	Gaussian model			Laplacian model		
			Mean value	Standard deviation	χ^2 Test	Median value	Mean deviation	χ^2 test
d (mm)	70	35	−16.9	5.9	60%	**−16.64**[b]	**4.4**	76%
$d_{f.l.}$ (mm)	35	17	−13.4	5.1	49%	−14.83	4.5	51%
r (mm)	52	26	**12.8**	**3.6**	92%	12.7	2.9	60%
δ (mm)	56	12	**21.4**	**3.8**	89%	22.1	3.1	57%
$\delta_{f.l.}$ (mm)	28	6	21.9	2.2	69%	22.3	1.7	69%
γ_1 (deg)	68	68	37.7	4.4	35%	**37.5**	**3.2**	61%
θ_1 (deg)	67	67	**40.3**	**3.6**	25%	40.	2.9	17%
γ_2 (deg)	54	54	**39**	**3.2**	67%	39.1	2.4	13%
θ_2 (deg)	52	52	41.6	3.8	44%	**41.8**	**3.0**	82%
α_f (deg)	70	10	6.6	0.5	97%	6.5	0.4	92%

[a] See Appendix 1.
[b] Figures in boldface indicate the most appropriate model shown graphically in Fig. 11.

range for all values. This evaluation can be pushed further by assuming momentarily that all observables fit a Gaussian distribution. Then, the internal variability σ_i of an observable within the cross-sections of an experiment should be close to its variability σ_t across all the data set. To illustrate this condition, consider a favorable case for the dips θ_1 and γ_1 (Fig. 7). All data are well mixed and do not cluster, mean values are well grouped and standard deviations for each experiment are close to the standard deviation for the whole data set. In addition, the standard deviation of the mean values (thick solid ellipse) is substantially smaller than

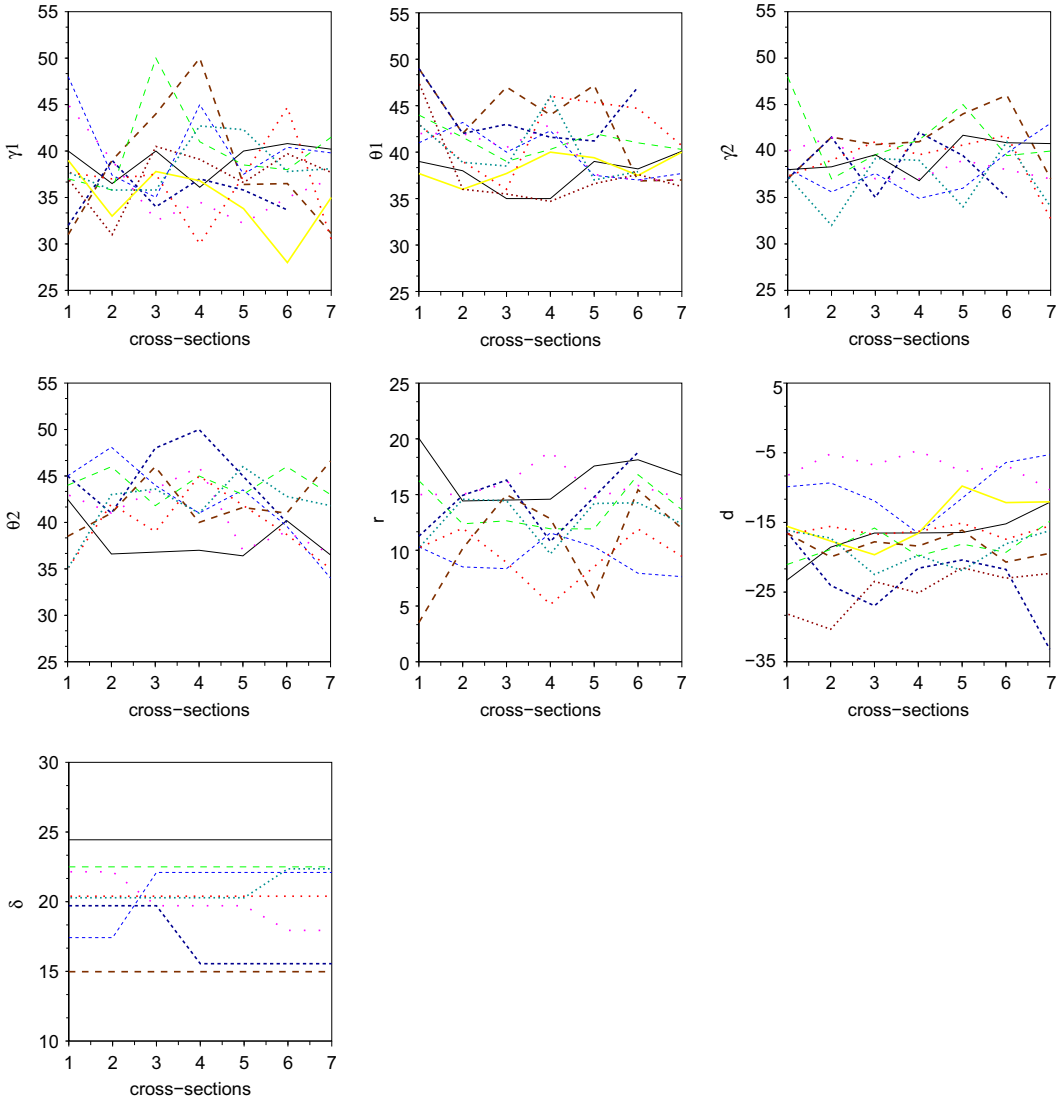

Fig. 6. Graphs of the measurements of each observable as functions of the cross-section number (Fig. 2c), in experiments of prototypes 1A and 1B. One color and line style per experiment.

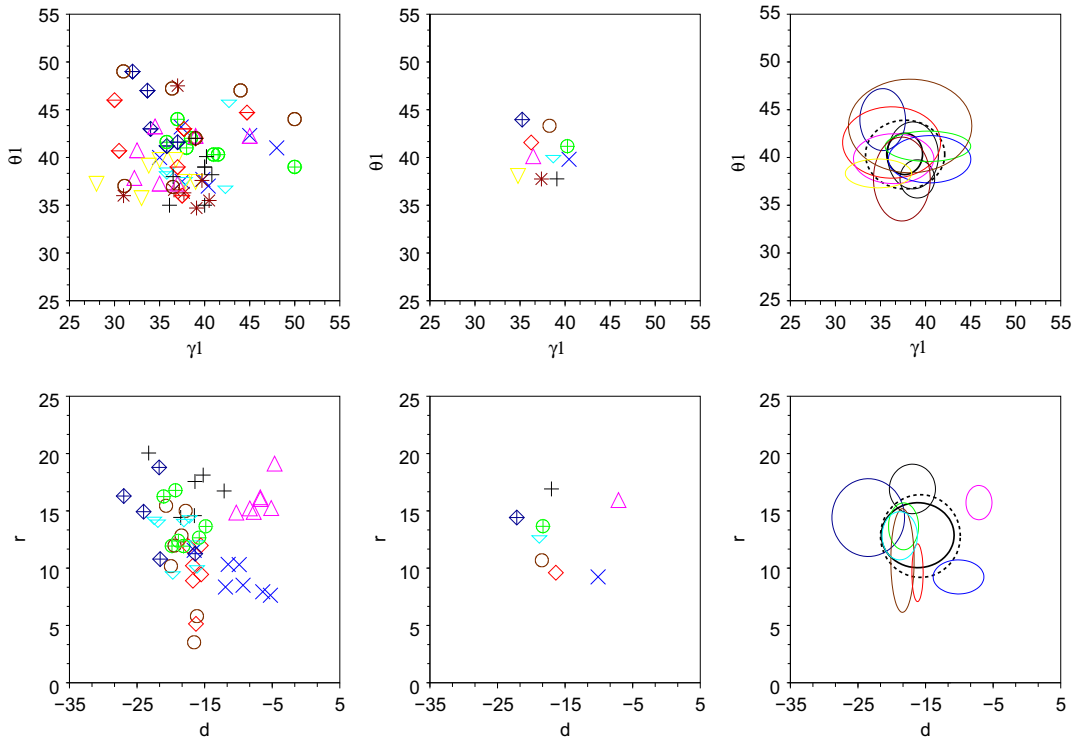

Fig. 7. Comparison of the variability of the measurements within each experiment versus the variability for the whole data set. Top: measurements of γ_1 and θ_1; bottom: measurements of d and r. Left column: all data with a different symbol and color for each experiment (same color code as in Fig. 6). Central column: mean values for each experiment. Right column: Ellipses with semi-axes lengths equal to the standard deviation for each experiment (thin solid colored lines), for the mean values shown in the central column (thick solid line), and for the whole data set (thick dashed line).

that of the whole data set (thick dashed ellipse), confirming the statistical convergence of the mean values of each experiment. This outcome means that the cross-sections provide an efficient sampling of these two observables.

For conciseness, we do not show these graphs for the 21 pairs of observables. Instead, we now show an unfavorable case, with d and r (recall that we could measure r in only eight experiments). Symbols for particular experiments cluster within the data field for all experiments, yet six means do cluster, and the results of particular experiments yield generally smaller deviations than for the entire data set. Furthermore, standard deviations of the whole data set (thick dashed) and of the mean values of each experiment (thick solid) are very close, indicating a slow statistical convergence of the cross-sections, i.e., a less efficient sampling of d and r, than of θ_1 and γ_1.

Based on the analysis of the angular observables, we conclude that the cross-sections separated by 20 mm, at roughly twice the frontal layer thickness $h = 9$ mm, can be interpreted as independent experiments when measuring fault dips. In contrast, we conclude that truly independent measurements of d and r occur only every second cross-section (i.e., for cross-sections separated by 40 mm, or $4.5 \times h$). The lifetime of the first forethrust, measured by δ, requires a much larger distance between cross-sections, around 20 cm (i.e., $22 \times h$). These conclusions will be used in section 4 to set the number N of independent measurements in the χ^2 tests for the fit of the data with Gaussian and Laplacian distributions (Table 2, column 2; Appendix 1).

A physical interpretation of these conclusions suggest that fault dips are dependent mainly on the intrinsic properties of the sand and of the initial state of packing, which are two particularly well-controlled steps in our experiments (same batch of sand, and use of a distributor). Fault locations would be more sensitive to less well-controlled experimental steps such as basal and lateral friction, or regularity of the initial sand pack, because these variations act indifferently on all cross-sections of an experiment, but differently between experiments. Finally, the nearly constant values of δ in each experiment indicates again that the second forethrust develops very rapidly along strike as shortening progresses. This outcome is reminiscent of studies for real fault displacement profiles that suggest a fast lateral growth of fault planes with respect to slip accumulation.

3.3. Identification of biases

3.3.1. Friction on the lateral walls

The effect of friction on the lateral walls is examined by comparing experimental results from lateral walls in configuration A to those in configuration B (prototypes 1A and 1B, Fig. 2, Table 1). The effects of friction are visually obvious for the edges of the deforming sand body because of the structural curvature (Fig. 2), so we focus on the seven central sections for each experiment. For these sections, d is plotted versus the other six geometric observables and the two data populations of each set of experiments for a prototype configuration are tested for similarity using a χ^2 test (Fig. 8 and appendix A2 for implementation). Classically in statistical works, the two distributions are considered similar for $Q \geq 5\%$. Clearly, d, θ_1, γ_2, θ_2, and r pass the test and for the central sections they demonstrate independence from the friction on the lateral walls. γ_1 fails the test ($Q = 0.36\%$) because of a mean value slightly higher (about 3°) in configuration A. Also, for these six geometric observables, the mean values with standard deviations overlap, which supports the conclusion that their measured characteristics in these central sections are independent of the experimental configuration and hence, the effects of friction on the wall.

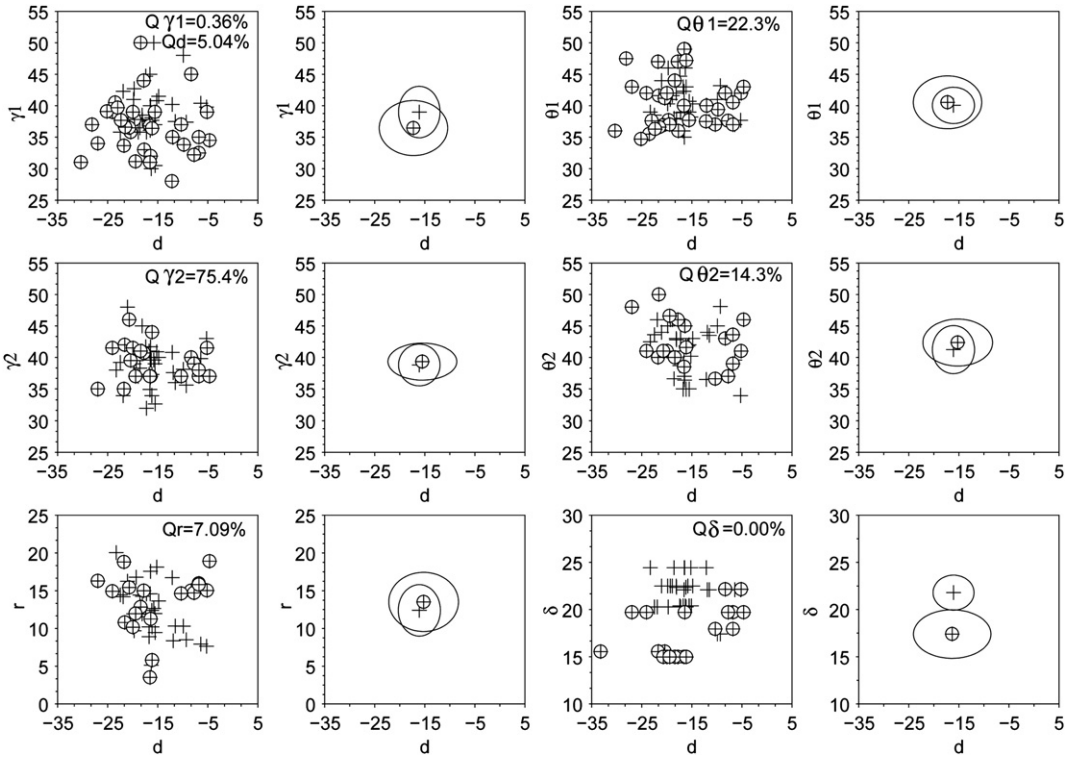

Fig. 8. First and third columns of graphs: measurements of six geometric observables versus the seventh geometric observable, d, in experiments of prototype 1A (crosses) and 1B (circled crosses). The Q values in the top right corners are results of χ^2 tests on the similarity of the two distributions. Second and fourth columns of graphs: mean values and standard deviations for the observables with same symbols.

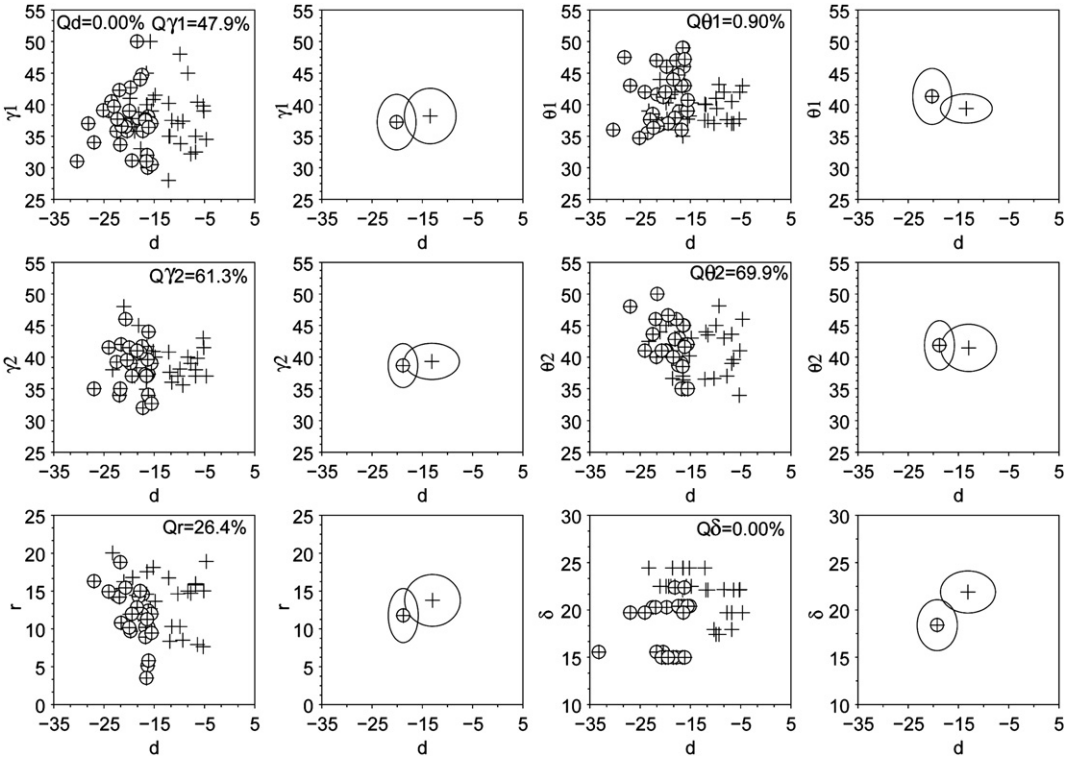

Fig. 9. First and third columns of graphs: measurements of six geometric observables versus the seventh geometric observable, d, in experiments using the discontinued lubricant (crosses) and the new lubricant (circled crosses). The Q values in the top right corners are results of χ^2 tests on the similarity of the two distributions. Second and fourth columns of graphs: mean values and standard deviations for the observables with same symbols.

δ shows strong sensitivity to the prototype configuration with data populations not fully overlapping and mean values with standard deviation not overlapping (Fig. 8). However, δ could not be measured for two 1B experiments, so δ would necessarily be greater than 25 mm, which would create more overlap between the results for 1A and 1B configurations.

3.3.2. Change of lubricant

Considering the two lubricants, the data was again separated into two populations: discontinued lubricant and current lubricant. Comparison of the two populations shows that, for γ_1, γ_2, θ_2, r, lubricant choice did not create a bias, and the populations are graphically and statistically similar (Fig. 9). For θ_1, the populations of the discontinued lubricant (crosses) display slightly lower mean and dispersion values, implying that this geometric observable is slightly biased by the lubricant choice. d and δ have higher values for the older (discontinued) lubricant, so a bias occurs because the older lubricant is more efficient when moving sand against glass.

3.3.3. Asymmetry of the box or sand pack

The last test concerns a possible asymmetry of the results across the plane cutting the box in two identical halfs, i.e. passing through the central cross-section (numbered 4 in Fig. 2c). The data, not shown for conciseness, exhibit similarity between the population for each half for each of the seven geometric observables. Therefore, asymmetries in the shape or density of the initial sand pack, or in the movement of the back wall with respect to the center plane do not introduce any meaningful bias.

3.3.4. Finite length of the box

The force sustained by the front wall was negligible in all experimental runs. This observation proves that the shortening force at the back wall was fully balanced in the frontal flat layer by basal and lateral friction, and thus that the finite length of the box did not affect results.

3.3.5. Example of a biased experiment

The effects of the biases can be demonstrated by using a slightly different prototype (Table 1, prototype 3). The sand layer is now much thicker than before, increasing contact with the lateral walls. The walls of the box were also not lubricated. Thrusts were curved in top view across the entire experiment and the front wall experienced compression (Fig. 10). These outcomes indicate that the box was too narrow with respect to the domain of frictional effects from the lateral walls and that the box needed to be longer to achieve a force balance within the model.

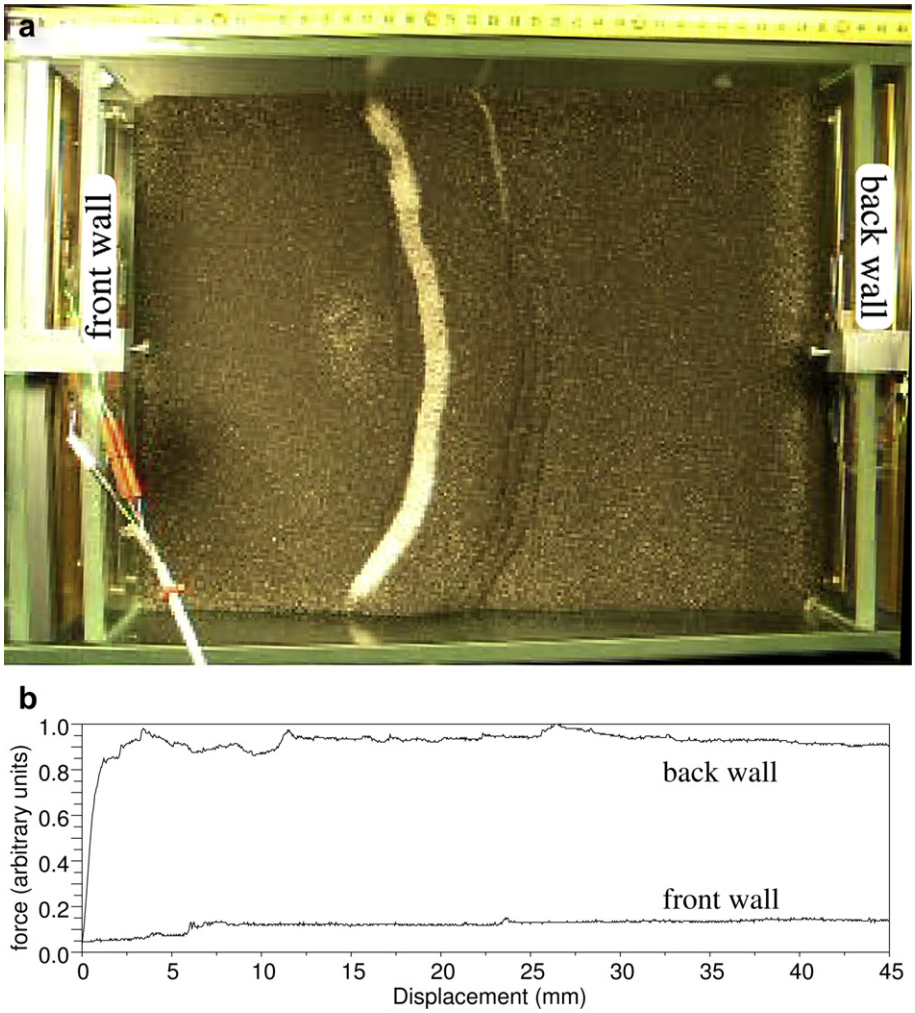

Fig. 10. Example of an experiment with two biases (prototype 3, Table 1). The top view of the final state in a) shows that shear stresses due to friction on the lateral walls curved the thrusts throughout the box width. The forces measured at back and front walls during shortening in b) show that some compressive stress has been transmitted to the front wall.

4. Statistical modeling of the observables

The statistical modeling aims at replacing the measurement populations by theoretical distributions which are amenable to theoretical treatment or to quantitative comparisons with e.g. numerical predictions of the experiments. Two classical monomodal distributions are likely adapted to our data: Gaussian and Laplacian. The difference is that in the Gaussian distribution the data are mostly gathered around their mean value whereas the Laplacian distribution allows them to be further away from their median value. χ^2 tests to measure the fit between the two theoretical distributions and the data are all above 10% and therefore, positive (Table 2, and implementation in Appendix 1). However, r, δ, θ_1, and γ_2 exhibit higher percentages with the Gaussian distribution. They are therefore best described by a Gaussian distribution whereas d, γ_1, and θ_2, best fit a Laplacian distribution (values in bold face). Data populations were graphically compared to the chosen theoretical distribution, and overall, the data populations do match the theoretical distributions with single modes (Fig. 11). Departures from the theoretical distributions are due to asymmetry (θ_1, γ_1, r), and excessive (d), or depressed (θ_1, θ_2, γ_2), central peak of the histograms. However they are not statistically significant because of the limited number of independent measurements. The distributions of d and δ using the first lubricant (lines $d_{f.l.}$ and $\delta_{f.l.}$, Table 2) display slightly greater mean or median values, interpreted, again, as the result of a better lubrification by the older product. The value of the surface slope α_f at the end of the experiment is slightly greater than at the beginning: $6.6^\circ \pm 0.5^\circ$.

Fig. 12 illustrates the uncertainty for the geometric observables at the onset of the second thrust-fault pair. The shading for the uncertainty is based on the standard deviation for the data populations matched to Gaussian distributions and for the mean deviation for populations matched to Laplacian distributions. This schematic illustration provides guidance for future experiments using these configurations (prototypes) in terms of the expected values for the geometric observables.

4.1. Modelling of the total shortening force

Measurements for 13 experiments (six Prototype 1A and seven 1B) were converted to Newtons by calibration of the strain gauges, which behaved in a linear elastic manner. The measurements were then processed using a 3 mm Hamming filter, so that each measurement was replaced by the average of the measurements located at $\pm$ 3 mm along the curve. The linear increase from zero to one mm displacement corresponds to the elastic deformation of the strain gauges. After reaching a broad maximum, five curves stabilize around a value for the normalized force of 0.85, seven curves stabilize around 1.05, and one curve stands alone around 1.35 (Fig. 13). We are not able to explain these variations in normalized force by the effect of the configuration of the lateral walls, nor by the change of lubricant, because comparisons of the data for each case do not show any differences. Slight changes in the tension of the gauges during the filling of the box with sand, resetting their zero level, is the most likely explanation of these variations.

Usually, localisation of deformation leading to the formation of thrusts is correlated to force variations (Krantz, 1991; Nieuwland et al., 2000; Lohrmann et al., 2003). Here, the correlation is very weak, or absent. This poor correlation is due to the thickness of the sand being very small compared to its horizontal dimensions. This setup was chosen to minimize the effect of lateral friction. With the same apparatus, stronger force variations were observed with thicker sand samples (Fig. 10). We believe therefore that the flatness of the curves of Fig. 13 reflects the true evolution of the force at the back wall. In the absence of a satisfactory correction procedure to guarantee the same zero level for each curve, we kept all data and compute a single mean curve with standard deviation (Fig. 13). Force measurements should prove to be a useful additional information when reproducing experiments by numerical methods or mechanical analysis.

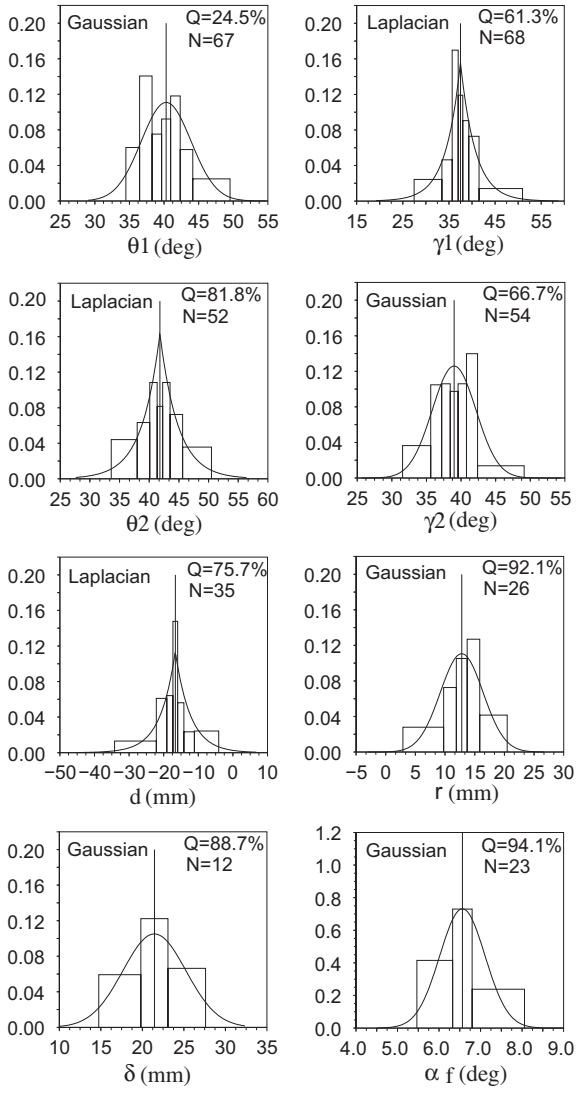

Fig. 11. Histograms of the measurements for each observable and the final wedge slope α_f, with theoretical distribution (Gaussian or Laplacian). Rectangular boxes– histograms, solid lines– theoretical distribution, vertical bars– mean value (Gaussian case), or median value (Laplacian case). The Q values in the top right corner of each graph indicate the confidence probability that the experimental distribution fits the theoretical one, according to a χ^2 test, and N indicates the number of independent measurements used to perform the test (see Appendix 1).

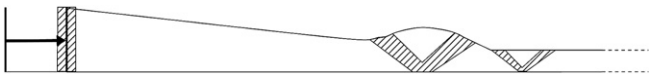

Fig. 12. Schematic cross-section for the experiments of prototypes 1A and 1B at the onset of the second thrusting event prior to the completion of experiments. The arrow shows the shortening and the shaded area illustrates the uncertainty for this shortening magnitude. The uncertainty for the positions and dips of the backthrusts and forethrusts is also shown with shading.

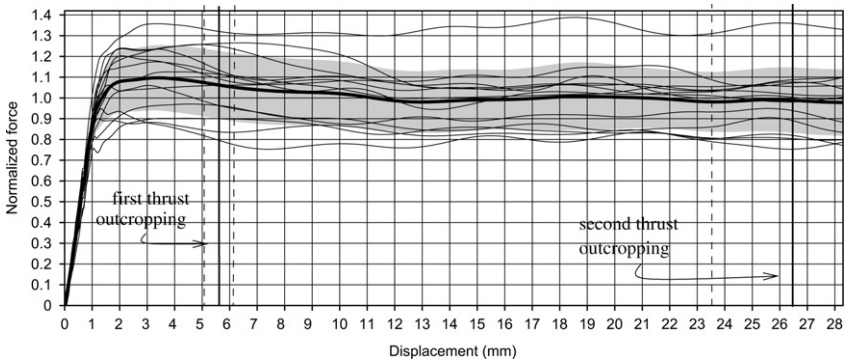

Fig. 13. Force sustained by the back wall during shortening. Thin solid lines– measurements of individual experiments, thick blak line– mean value of all experiments, grey area– standard deviation from the mean value.

Table 3
Statistical models for observables of prototype 2 experiments.

	Nbr. of meas.	Nbr. *N* of indep. meas.	Gaussian model			Laplacian model		
			Mean value	Standard deviation	st. dev. from exp. 1	Median value	Mean deviation	mean dev. from exp. 1
d (mm)	28	14	−3.9	7	5.9	−4.1	5.1	4.4
r (mm)	12	6	15.6	3	3.6	14.6	2.5	2.9
δ (mm)	28	4.8	23.9	1.3	3.8	23.9	0.9	3.1
γ_1 (deg)	24	24	39	6.5	4.4	39.7	5.4	3.2
θ_1 (deg)	24	24	38.2	4.9	3.6	37.1	3.8	2.9
γ_2 (deg)	14	14	39.9	3.5	3.2	37.7	2.9	2.4
θ_2 (deg)	12	12	41.9	3.4	3.8	42.8	2.6	3.0
α_f (deg)	28	4	7.5	0.5	0.5	7.5	0.35	0.4

4.2. Application to other experiments

To test whether the variations of the observables are valid for other experimental configurations, we consider experiments performed using the prototypes 2A and 2B (Table 1), which differ from experiments of prototypes 1A and 1B by the surface slope α, now at 7° instead of 6°, and the basal slope β, now at 3° instead of 0° (Fig. 3a). Each experiment 2A and 2B was repeated two times, using the same protocol and yielded a number of measurements for each observable (Table 3). From these measurements, we computed the central and dispersion estimators for each observable in the Gaussian and Laplacian cases (columns 3, 4, and 6, 7, of Table 3, respectively), and we reproduce the dispersion estimators of experiments 1 in columns 5 and 8, for comparison. Considering a graphical comparison of Prototype 2 data versus appropriate Gaussian or Laplacian distributions using Prototype 1 dispersion estimators (Fig. 14), visual comparison of these curves with the histograms allow us to conclude that the variations of d, γ_1, θ_1, and γ_2 are well represented by the error bars of experiments 1. χ^2 tests support these conclusions for γ_1 (Laplacian case) and θ_1 (both cases). The number of independent measurements is generally insufficient to rely on χ^2 tests for θ_2, r, δ, α_f ($N < 14$). Finally, although only 4 independent measurements exist for the prototype 2 experiments, it is worth noting that the surface slope α_f at the end of the experiments behaves as in experiments 1, where it is greater than the initial slope α by half a degree, with the same dispersion.

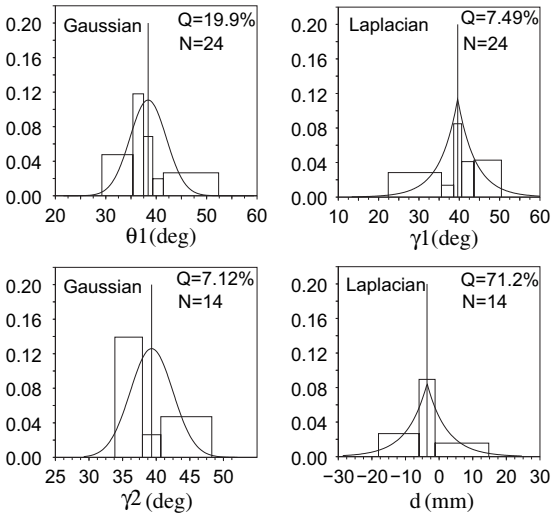

Fig. 14. Histograms for the measurements of four observables of experiments 2A and 2B. Rectangular boxes– histograms; solid lines– theoretical distribution, vertical bar– mean value (Gaussian case), or median value (Laplacian case). The theoretical distributions are constructed using the dispersion estimators of experiments of prototypes 1A and 1B (Table 3, columns 5 and 8 respectively).

5. Conclusion

We have presented generic experiments of the quasi-static growth by shortening in plane-strain conditions, of a sand wedge resting on a flat sand layer. Our goal is to cast the observed evolution of thrusting into a set of geometric and mechanical measurable quantities (or observables), and then to construct statistical distributions describing their variations. A specific experimental apparatus allowed us to produce rapidly uniform sand packs, reverse the sense of friction on the lateral walls, and measure the forces necessary to shorten the sand body. The main conclusions of the analysis and statistical modeling are:

- Gauss or Laplace distributions adequately model our measurements. Typical error bars for fault dips are around 3.3°, and around 3.8 mm for fault locations.

- Cross-sections can be considered as independent experiments (hypothesis of ergodicity) if they are separated by 20 mm when measuring fault dips, 40 mm for fault locations, and about 200 mm for the lifetime of the first thrusting event.
- The finite width of the box creates a bias due to shear stress on the lateral walls that disappears at a distance of about 80 mm from the walls. The finite length of the box did not influence the results.
- Error bars remain valid for experiments using slightly different setups (dipping basal plate and wider wedge taper angle).

It should be underlined that a change of material, protocol, or setup would change the figures determined here, but the statistical analyses can be used in any laboratory, in order to construct appropriate statistical models. These models allow us to cast the 3D experiments into a 2D cross-section with error bars. This is a necessary step towards a quantitative comparison with theoretical predictions.

Acknowledgements

Engineers Jean-Marc Siffre and Jean-Christian Colombier have brought a decisive expertise in the construction of the experimental apparatus. This apparatus was financed as part of a research agreement between the University of Cergy-Pontoise and the Institut Français du Pétrole. Most graphics and all statistical programs were done using the free software Scilab (www.scilab.org). We thank journal editor William Dunne for his detailed review of the text.

Appendix A1. χ^2 test on the fit between an experimental and a theoretical distribution

We have N independent measurements of an observable, and we would like to test whether they follow a given theoretical distribution. First, we build a histogram of the data with p bins (definition of the bins is treated below). Let N_i be the actual number of samples in bin i, and N_i^{th}, the number of samples in bin i according to the theoretical distribution. We then calculate the fit between the two histograms as (Press et al., 1992, equation (14.3.1))

$$D^2 = N\sum_{i=1}^{p}\frac{\left(f_i - f_i^{th}\right)^2}{f_i^{th}}, \tag{2}$$

where $f_i = N_i/N$, and $f_i^{\text{th}} = N_i^{\text{th}}/N$ and the result of the test is usually named Q and is one minus the image of D^2 by the cumulative distribution function of χ^2_{p-1} (this can be found in standard statistical libraries of usual softwares such as Matlab or Scilab). A Q value of one (obtained when $D^2 = 0$) indicates a perfect fit, while a value of zero indicates a complete discrepancy between the two distributions. These values are given in percentage in the graphs of Figs. 11 and 14. In practice, a fit is accepted for $Q \geq 10\%$, and rejected for $Q \leq 5\%$.

Choice of the number of independent measurements for each observable

The test is rather sensitive to the value of N (Table 2, column 2): the same histogram may pass the test if N is small, and may be rejected if N is large. For d and r, we define N as half the number of actual measurements because we showed in Section 3.2 that the cross-sections were too closely spaced for these observables. Considering that only half of the measurements are independent when performing the tests is equivalent to having considered cross-sections separated by 40 mm instead of 20 mm. The spacing of 20 mm appeared satisfactory for all fault dips, and therefore N is the actual number of measurements for θ_1, γ_1, θ_2, γ_2. δ, the lifetime of the first thrust, could be measured in only eight experiments out of ten, and exhibited lateral variations in half of these eight experiments. We therefore define N for δ as $N = 8 + 8/2 = 12$. Finally, the surface slope of the wedge at the end of shortening, α_f, exhibited a constant value in all cross-sections of each of the ten experiments, and therefore $N = 10$ instead of 70. All these values are summarized in Table 2 for experiments 1A and 1B. Tests concerning experiments 2A and 2B relied on the same arguments (Table 3).

Definition of bins to construct histograms for a comparison with a theoretical distribution

Assume we have N independent measurements of an observable and want to construct a histogram of their values for comparison with a theoretical distribution. Typically, the number of bins $p = 9$ if $N \geq 45$, $p = 5$ for $25 \leq N \leq 45$, and $p = 3$ for $N \leq 25$, so that at least five samples occur in each bin, otherwise the statistical test is not reliable. The $p + 1$ quantiles that define the limits of the p bins are defined as follows. First, compute from the data the central and dispersion estimators (e.g., mean value and standard deviation for the Gauss distribution), which define the theoretical probability density function of the data. Then, compute by integration the corresponding cumulative density function (CDF) which has values between 0 and 1. The quantiles are the values such that their image by the CDF are regularly spaced between 0 and 1. The first and last quantiles are however particular cases that depend on data. They are defined slightly below and above the minimum and maximum measurement, respectively.

Appendix A2. χ^2 test on the similarity between two experimental distributions

Assume we have N measurements of an observable, and we separate these into two groups of N_R and N_S measurements, respectively. We would like to decide whether these two distributions are similar, to determine for example the effect of a change of experimental setup on our measurements. We first build a histogram of the data with p bins (definition of bins is treated below). Let R_i and S_i be the numbers of measurements falling into bin i in each group respectively. We then calculate (Press et al., 1992, equation (14.3.3))

$$D^2 = \sum_{i=1}^{p}\frac{\left(\sqrt{N_S/N_R}R_i - \sqrt{N_R/N_S}S_i\right)^2}{R_i + S_i}, \tag{3}$$

and the result of the test is Q in percentage, defined and interpreted as in Appendix 1.

Definition of bins to construct two histograms to be compared

We gather the two groups into a single group of N measurements. We order the samples, say in increasing order, and we group them in p bins such that the same percentage of measurements occurs in each bin (the amounts may also vary for each bin, if desired). Then we compute the limits of the bins as follows: for limits between two bins, we define the limit as the middle between the highest sample of the lower bin and the lowest sample of the higher bin. The minimum limit is defined as the lowest sample minus the mean distance between samples of the lowest bin. Likewise, the maximum limit is defined as the highest sample plus

the mean distance between samples of the highest bin. The test is then performed using the two groups, N_R and N_S, separately.

References

Bonnet, C., Malavieille, J., Mosar, J., 2007. Interactions between tectonics, erosion, and sedimentation during the recent evolution of the Alpine orogen: analogue modeling insights. Tectonics 26, 6.

Buiter, S.J.H., Babeyko, A.Y., Ellis, S., Gerya, T.V., Kaus, B.J.P., Kellner, A., Schreurs, G., Yamada, Y., 2006. The numerical sandbox: comparison of model results for a shortening and an extension experiment. In: Buiter, S.J.H., Schreurs, G. (Eds.), Analogue and numerical modelling of crustal-scale processes. Geological Society, London, Special Publication, 253, pp. 29–64.

Cadell, H.M., 1888. Experimental researches in mountain building. Royal Society of Edinburgh Transactions 35, 337–360.

Costa, E., Vendeville, B., 2004. Experimental insights on the geometry and kinematics of fold-and-thrust belts above weak, viscous evaporitic décollement: reply to comments by Hemin Koyi and James Cotton. Journal of Structural Geology 26, 2139–2143.

Cubas, N., Leroy, Y.M., Maillot, B., 2008. Prediction of thrusting sequences in accretionnary wedges. Journal of Geophysical Research 113, B12412. doi:10.1029/2008JB005717.

Crook, A.J.L., Wilson, S.M., Yu, J.G., Owen, D.R.J., 2006. Predictive modelling of structure evolution in sandbox experiments. Journal of Structural Geology 28, 729–744.

Davis, D., Suppe, J., Dahlen, F.A., 1983. Mechanics of fold-and-thrust belts and accretionary wedges. Journal of Geophysical Research B2, 1153–1172.

Del Castello, M., Cooke, M.L., 2007. Underthrusting-accretion cycle: work budget as revealed by the boundary element method. Journal of Geophysical Research (Solid Earth). doi:10.1029/2007JB004997.

Egholm, D.L., Sandiford, M., Clausen, O.R., Nielsen, S.B., 2007. A new strategy for discrete element numerical models: 2. Sandbox applications. Journal of Geophysical Research 112, B0524.

Ellis, S., Schreurs, G., Panien, M., 2004. Comparisons between analogue and numerical models of thrust wedge development. Journal of Structural Geology 26 (9), 1659–1675.

Hardy, S., McClay, K., Muñoz, J.A., 2009. Deformation and fault activity in space and time in high-resolution numerical models of doubly vergent thrust wedges. Marine and Petroleum Geology 26, 232–248. doi:10.1016/j.marpetgeo.2007.12.003.

King Hubert, M., 1951. Mechanical basis for certain familiar geologic structures. Bulletin of the Geological Society of America 62, 355–372.

Klinkmüller, M., Rosenau, M., Boutelier, D., Kemnitz, H., Schreurs, G., 2008. Properties benchmark of granular and viscous analogue materials. Extended abstract, Bollettino di Geofisica, vol. 49, No. 2 supplement. In: GeoMod2008 extended abstracts, Firenze, 21–24 September 2008.

Koyi, H.A., 1997. Analogue modelling: from a qualitative to a quantitative technique, a historical outline. Journal of Petroleum Geology 20 (2), 223–238.

Koyi, H.A., Cotton, J., 2004. Experimental insights on the geometry and kinematics of fold-and-thrust belts above weak, viscous evaporitic décollement; a discussion. Journal of Structural Geology 26, 2139–2143.

Krantz, R.W., 1991. Measurements of friction coefficients and cohesion for faulting and fault reactivation in laboratory models using sand and sand mixtures. Tectonophysics 188, 203–207.

Lohrmann, J., Kukowski, N., Adam, J., Oncken, O., 2003. The impact of analogue material properties on the geometry, kinematics, and dynamics of convergent sand wedges. Journal of Structural Geology 25, 1691–1711.

Maillot, B., Barnes, C., Mengus, J.-M., Daniel, J.-M., 2007. Constraints on friction coefficients by an inverse analysis of sand box thrust dips. Journal of Structural Geology 29, 117–128.

Maillot, B., Koyi, H.A., 2006. Thrust dip and thrust refraction in fault-bend-folds: analogue models and theoretical predictions. Journal of Structural Geology 28, 36–49.

Mulugeta, G., Koyi, H., 1987. Three-dimensional geometry and kinematics of experimental piggyback thrusting. Geology 15, 1052–1056.

Mulugeta, G., Koyi, H., 1992. Episodic accretion and strain partitioning in a model sand wedge. Tectonophysics 202, 319–333.

Nieuwland, D.A., Urai, J.l., Knoop, M., 2000. In-situ stress measurements in model experiments of tectonic faulting. In: Lehner, F.K., Urai, J.L. (Eds.), Aspects of Tectonic Faulting. Springer-Verlag, pp. 155–167.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in Fortran, the Art of Scientific Computing, second ed.. Cambridge University Press, Cambridge, England.

Saltzer, S.D., 1992. Boundary-conditions in sandbox models of crustal extension – an analysis using distinct elements. Tectonophysics 215 (3–4), 349–362.

Schreurs, G., Buiter, S.J.H., Boutelier, D., Corti, G., Costa, E., Cruden, A.R., Daniel, J.-M., Hoth, S., Koyi, H.A., Kukowski, N., Lohrmann, J., Ravaglia, A., Schlische, R.W., Oliver Withjack, M., Yamada, Y., Cavozzi, C., Delventisette, C., Elder Brady, J.A., Hoffmann-Rothe, A., Mengus, J.M., Montanari, D., Nilforoushan, F., 2006. Analogue benchmarks of shortening and extension experiments. In: Buiter, S.J.H., Schreurs, G. (Eds.), Analogue and Numerical Modelling of Crustal-Scale Processes. Geological Society, London, Special Publication 253, pp. 1–27.

Seyferth, M., Henk, A., 2006. A numerical sandbox: high-resolution distinct element models of half-graben formation. International Journal of Earth Sciences 95 (2), 189–203.

Vendeville, B.C., 2007. The 3-D Nature of Stress Fields in Physical Experiments and Its Impact on Models Overall Evolution. Solicited Oral Contribution. In: EGU General Assembly, Vienna, 15–20 April, 2007.

Wygal, R.J., 1963. Construction of models that simulate oil reservoirs. Society of Petroleum Engineers Journal, 281–286.