Building a Geological Reference Platform Using Sequence Stratigraphy Combined with Geostatistical Tools

Bernard Bourgine, Éric Lasseur, Aurélien Leynet, Guillaume Badinier, Carole Ortega, Benoît Issautier, and Valentin Bouchet

Abstract This paper presents a methodology that is currently tested at the French geological survey in order to validate drill holes interpretation. Validated drill holes are intended to be included in the future French geological reference platform which is under construction. To validate drill holes, a first subset of high-quality holes is selected. This data is interpreted in terms of geology and a geostatistical analysis is performed. A 3D geological model is built to assess the overall geological consistency. Then the rest of the drill holes is progressively and iteratively validated by geostatistical cross validation. As several thousands of drill holes are to be validated, specific software and workflows have been developed and are presented here.

1 The French Geological Reference Platform

The objective of this work is the setup of a methodology for the validation of drill holes that will be included in the French geological reference platform.

The French geological reference platform (Fig. 1) is one of the major scientific programs of the BRGM (BRGM is the French geological survey). It aims at delivering 3D validated geological data to public institutes, to the scientific community or to private companies, as well as to every citizen.

The data to be validated is of various types: geological maps, well logs, field data, seismic profiles, etc.

B. Bourgine (✉) • É. Lasseur • A. Leynet • G. Badinier • C. Ortega • B. Issautier
BRGM (French Geological Survey), 36009, 45060 Orléans Cedex 2, France
e-mail: b.bourgine@brgm.fr; e.lasseur@brgm.fr; a.leynet@brgm.fr; g.badinier@brgm.fr; c.ortega@brgm.fr; b.issautier@brgm.fr

V. Bouchet
ISTO, Orléans, France
e-mail: valentin.bouchet1@gmail.com

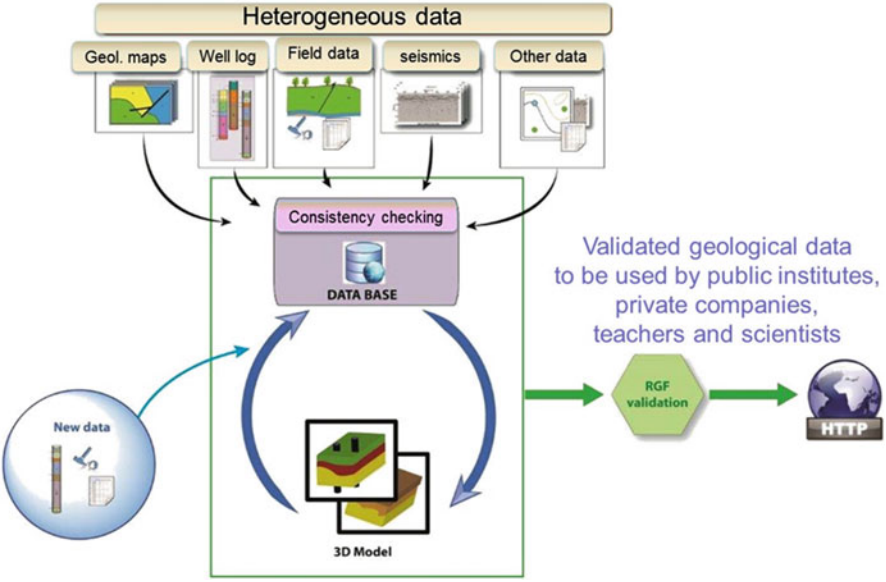


Fig. 1 The French geological reference platform

It is heterogeneous data, stored presently in independent databases. Moreover, it has often not been homogenized, verified, or compared to other data. For this reason this data cannot always be used directly and is not always reliable.

In the geological reference platform that will be built, every type of data will be stored in a normalized database and will share the same data model.

This will allow data comparison and data checking and will facilitate data interpretation and data use.

Regarding data validation, several levels of validation will be defined, but one of the key components will be the validation through the construction of 3D geological models or other types of numerical models such as flow simulation or geophysical inversion.

For new data acquired later, tools will be developed in order to check its consistency toward other existing data and 3D models.

In the end, validated data and models will be made accessible on the Internet.

2 Definition and Objective of Drill Holes Validation

Drill holes are one of the most important subsurface data to be validated in the RGF (french geological reference platform). These drill holes are currently stored in the French national drill holes database (BSS), which actually includes more than 800,000 drill holes. For most of the drill holes, the only provided information is the

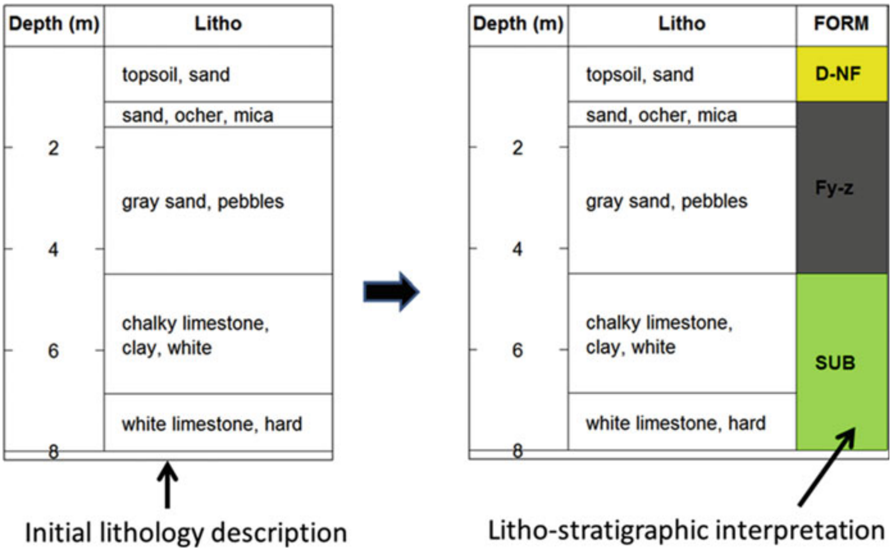


Fig. 2 First litho-stratigraphic interpretation from initial lithology description of drill holes (*D-NF* marine dunes, *Fy-z* quaternary alluviums, *SUB* bedrock)

location of the hole, as well as the description of the lithology encountered at different depths along the hole.

This information in itself is not sufficient to enable direct use of the drill holes to build a geological model. For this it is necessary at least to associate each lithology encountered along the hole with a known geological formation, a formation age, or a geological body.

This is why a preliminary litho-stratigraphic interpretation was made in the past, for a subset of 10 % of the database, approximately 90,000 holes (Fig. 2). This first level of interpretation includes the definition of lexicons concerning standardized formation names, their age, and their lithology, as well as their graphical representation.

However, the quality of this preliminary interpretation can range from very good to very low, depending on the quality of the original drill holes description.

Moreover, the drill holes were often interpreted independently, without taking into account other drill holes around, geological maps, or other information.

This is why a next step of the process is planned in the framework of the RGF. The objective is to reprocess the 90,000 drill holes in order to validate or not the preliminary interpretation in terms of litho-stratigraphy and to record and validate the main geological interfaces crossed by drill holes.

As many drill holes are to be processed and as each drill hole can intersect frequently 10–20 interfaces, it is of course necessary to automatize the work as much as possible, in order not to spend too much time and money.

3 Overview of Drill Holes Validation Methodology

Basically, the methodology is divided into two steps (Fig. 3):

1. A reference set (set *A*) is created using a limited number of high-quality drill holes.
2. Other drill holes of lower quality (set *B*) are validated progressively by comparison to set *A* and are iteratively added to set *A*.

A geostatistical analysis of set *A* data enables data characterization as well as a quantification of uncertainty and helps validating interpretation.

In the first step, the reference set (set *A*) is created by selecting a loose network of high-quality drill holes, i.e., holes owning a well log, generally a gamma ray. These well logs are correlated using sequence stratigraphy (Homewood et al. 1999), which allows an accurate litho-stratigraphic interpretation. A geostatistical analysis of interpreted data is performed. The objective of this analysis is to spatially characterize this data, detect potential errors (by cross validation), and establish the variogram model then used. 3D models are built in order to quantify uncertainty and validate the set *A* by checking the overall consistency of geological bodies with respect to geological knowledge. Isopach maps derived from the model contribute to this validation procedure.

During step 2 all other drill holes (of lower and unknown quality) are checked. A geostatistical cross validation technique is used to check their consistency as compared to the reference set *A*. Drill holes that are in good agreement with reference data are considered as validated and added to the reference set. Generally, the geostatistical model itself is not challenged because it is supposed to have been validated at step 1. However, in case of high inconsistency, it can be reviewed. Then the process is iterated to look for other drill holes to validate. 3D models are updated consequently and results verified (Fig. 3).

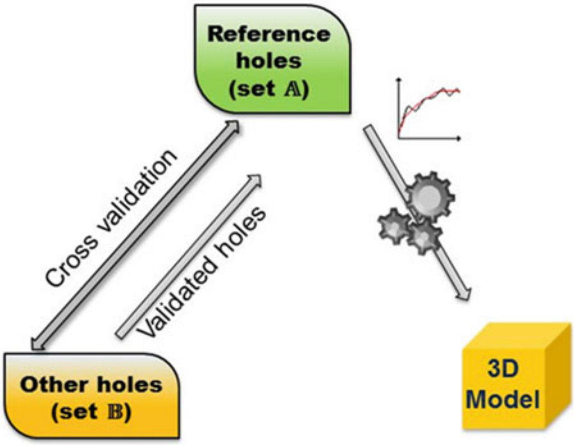
4 Application of the Methodology on Real Data

The methodology of validation of drill holes will be now presented in more details and illustrated on two real datasets, one in the Paris basin and the other in the North Aquitanian basin (France).

4.1 First Step: Analysis of the Reference Set

The first example shown in Fig. 4 is a 120 * 80 km area around Paris. In this area, approximately 7000 drill holes are to be validated.

Fig. 3 Overall methodology



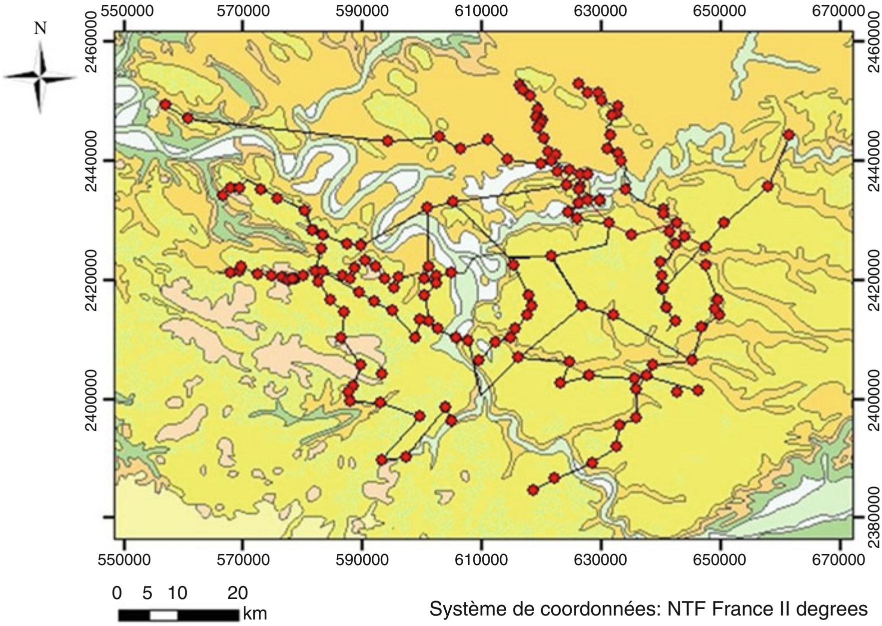


Fig. 4 Dataset in the tertiary basin around Paris, France

Among them, 168 reference drill holes with a gamma ray are chosen for the reference set (2.5 %).

These reference drill holes are grouped along vertical transects (Fig. 4) that are aligned along the major geological axes. Drill holes and well logging are displayed along these transects (e.g., Fig. 5 in the Aquitanian basin). The geologist can interpret well logs using sequence stratigraphy concepts. Basically, the geologist

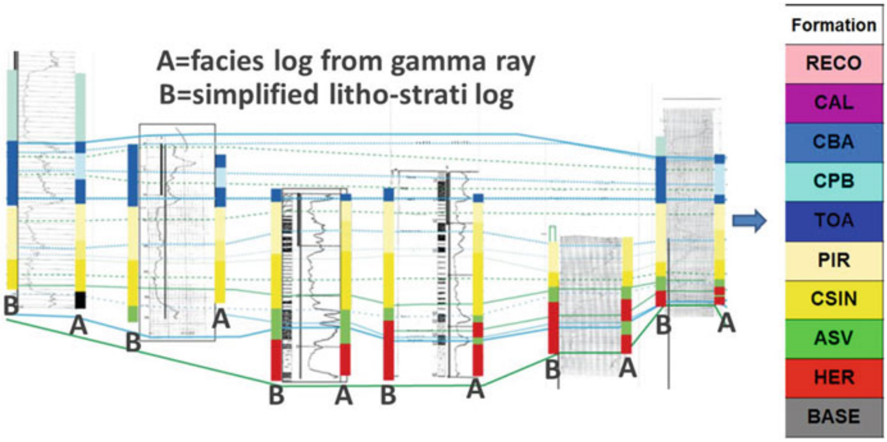


Fig. 5 Interpretation of drill holes using gamma ray logging and sequence stratigraphy concepts. Example in the Aquitanian basin

has to identify surfaces or time lines that correspond to depositional context change. This allows restoring correctly the geometry of sedimentary bodies. This will also make possible facies simulation in each geological body.

Once the well logs are interpreted in terms of sequence stratigraphy, it is possible to define a litho-stratigraphic pile along the wells (Fig. 5). The litho-stratigraphic pile interpretation defined here (shown at the left of each gamma ray (Fig. 5)) is a simplification of sequence stratigraphy interpretation (shown at the right of each gamma ray on Fig. 5), where several different facies are grouped together between two "major" time lines.

A geostatistical analysis of reference drill holes is then performed. The studied variable is the elevation of the top or base of each formation. It can also be the formation thickness.

For each formation, a "formation status map" is drawn. It shows if the formation is present or absent and if the drill hole has intersected only the top of formation, the base of formation, or both top and base (Fig. 6).

For example, in Fig. 6 the green "#" symbols correspond to a formation gap. When isolated and surrounded by holes intersecting the formation, it can indicate an error of interpretation or a coordinate error. Otherwise, these "#" symbols show where the formation has not deposited and help the geologist to check his interpretation.

From a geostatistical point of view, the variograms of tops and bottoms are computed and fitted. Generally, a polynomial drift (linear or quadratic) is considered, and the fitted variogram is that of the residual after removal of a global drift (fitted by least squares to all data – which is an approximation of the true residual). Figure 7 shows an example of variogram of the residual computed from high-quality data. The error in the interpretation of logging is of the order of few

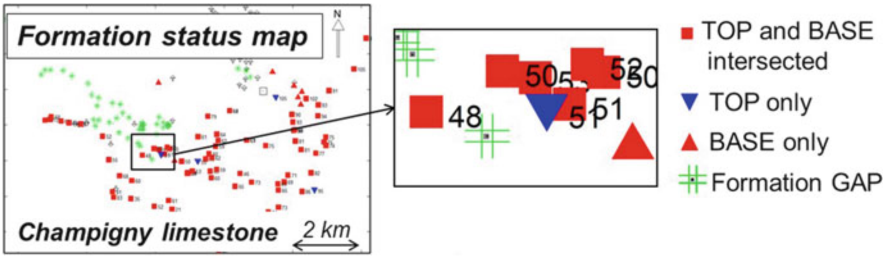


Fig. 6 Formation status map. Example for Champigny limestone. Paris basin

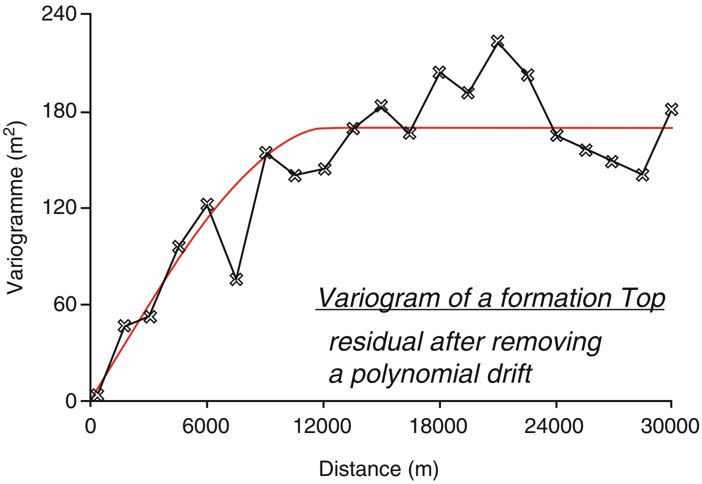


Fig. 7 Example of the variogram of the top of a formation (residual after removing a polynomial drift)

centimeters, and the nugget effect can be neglected. It would not be the case with destructive holes for which the identification of formation change, based on cuttings, can be affected by an error of 1–2 m.

Then a cross validation (based on kriging) is then performed. At this step, the coefficient of the drift function is computed locally by kriging using data of the neighborhood, so the local drift (and then the local residual) may differ from the global one used to fit the variogram of the residual. Ideally, the new residual should be recomputed and the process should be iterated as shown by Hengl et al. (2007). This is not done in our case. Consequently, there is a risk of error due to a nonoptimal geostatistical model. However, in practice the cross validation of the reference data gives very satisfactory results in terms of error and of normalized error, and outliers found in data can generally be attributed to errors in the data. In our case the approximation seems acceptable and the geostatistical model is validated.

During this cross validation, inequality constraints given by the drill holes that have not crossed or reached a given formation are taken into account. For example, if a drill hole does not reach a given formation, this gives an upper bound value for the elevation of the top of this formation. When the top of this formation is interpolated from other drill holes where it has been observed, the interpolated value at the location of the drill hole is compared to the upper bound. A warning is sent if the interpolated value is above the upper bound and data can be verified. Note that at this step a standard kriging is used instead of a "kriging taking into account inequalities" (Freulon and De Fouquet 1993; Abrahamsen and Benth 2001) because the objective is first to check data consistency and detect potential error (even in the inequality constraint themselves).

Other information like the DTM and the geological map are also taken into account. For example, the geological map, if correct, virtually gives infinity of inequality constraints: if at a given point the formation *A* is outcropping that gives constraints on the elevation of other geological formations.

The whole litho-stratigraphic pile has to be processed. If the litho-stratigraphic pile in the area includes 20 formations, this procedure has to be repeated for all the formations.

To make this work possible in a consistent and largely automated way, specific software has been developed by the BRGM: GDM MultiLayer (Geological Data Management, http://gdm.brgm.fr) (Bourgine et al. 2008; Bourgine 2015). This software and associated algorithms include the management of gaps due to erosion or formation pinching.

At last a preliminary 3D geological model is built using the reference dataset. This helps checking the overall interpretation and possibly to correct it.

4.2 Second Step: Validation of Remaining Drill Holes

Once the reference dataset is validated for all the formations, other drill holes (dataset *B*) are compared to the reference dataset.

For this, we again apply a cross validation technique. For example, if we consider the top of a formation, and if we denote *TA* as the top of the formation known from reference drill holes (set *A*) and *TB* the top of the same formation measured on drill holes dataset *B*, we estimate *TB* from drill holes belonging to dataset *A* and compare estimated value *TB** to the true value *TB*. Thus, we can compute the estimation error (*TB*-TB*) and the normalized error if we divide by the kriging standard deviation.

The *B* holes, where we do not know the true value *TB*, but an inequality constraint on *TB*, can also be tested.

At last the software we have developed displays automatic maps highlighting outliers and potential errors. By clicking on the map, the geologist can get vertical cross sections showing the potential anomaly.

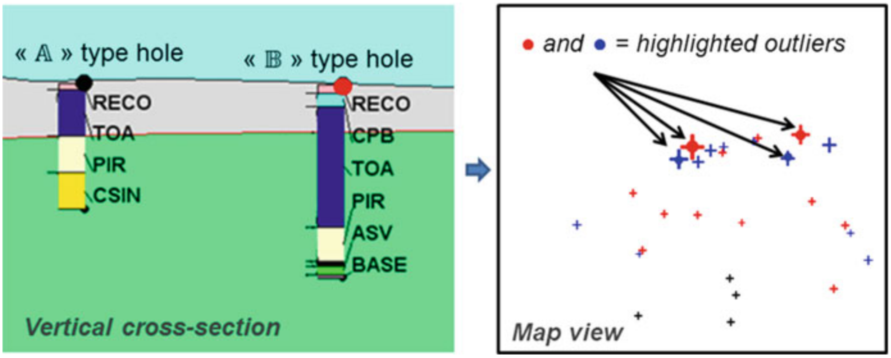


Fig. 8 Validation of B holes along a cross section. Example for the top of "TOA" formation (dark-blue formation along holes)

For example, on the cross section in Fig. 8 (data from the Aquitanian basin), the interpolation of the base of the dark-blue formation (named *TOA*) using reference holes *A* is not consistent with *B* hole where the base of dark-blue formation is intersected much lower. Either drill hole *B* is correct (e.g., in the case of a fault or a fold between the two holes) or drill hole *B* is not located here (not GPS-tagged drill hole, coordinate input error, database error) or has not been correctly interpreted. Typically, this hole has to be validated or corrected by manual check.

In fact all the *B* holes are not checked systematically, but only when necessary. In a first step, *B* holes which are consistent with reference holes and with the geological model are validated automatically. These new validated holes are then added to the reference set *A*, and the process is reiterated with nonvalidated holes remaining in set *B*. The iterative process ends when there are no more automatically validated holes in set *B*. The holes remaining in set *B* can then be checked and corrected manually. At each step, the variogram can (on demand, if necessary) be recomputed and remodeled, especially to improve fitting for the short-range components.

For the selection of automatically validated holes, we use the table shown in Fig. 9. Drill holes for which the estimation error is low and the normalized error is lower than 2 (green cell in Fig. 9) are validated automatically and added to the reference set *A*. The threshold between "low" and "high" estimation error is fixed arbitrarily and depends on geological context, as well as on the lithology contrast between successive formations, which can make the drill holes interpretation easy or difficult.

The threshold value of 2 for the normalized error corresponds to a 95 % confidence interval if we assume that the kriging error distribution is Gaussian. It is well known that kriging error is generally not Gaussian. However, we have verified that it is a reasonable approximation in our case.

		Estimation error	
		Low $\lvert T^*_{\mathbb{B}} - T_{\mathbb{B}} \rvert$	**High** $\lvert T^*_{\mathbb{B}} - T_{\mathbb{B}} \rvert$
Deviation to the model	$\lvert T^*_{\mathbb{B}} - T_{\mathbb{B}} \rvert / \sigma_K < 2$	hole $\mathbb{B}$ validated	hole $\mathbb{B}$ to be verified
	$\lvert T^*_{\mathbb{B}} - T_{\mathbb{B}} \rvert / \sigma_K > 2$	hole $\mathbb{B}$ to be verified	hole $\mathbb{B}$ to be verified

Fig. 9 Criterion for validating or not drill holes of type B

All the holes that are not automatically validated are kept for further checking, and the process is iterated using reference *A* holes plus the validated *B* holes, as soon as there is no new auto-validated *B* hole.

The *B* holes that are not automatically validated can be checked manually one by one. As it is not always possible to check all drill holes, we check preferentially drill holes that are located in critical zones, for example, that intersect important aquifers. Outliers of the cross validation are also checked because often they reveal a major error in the interpretation or in the data.

Another criterion is to select the next candidate for validation in areas where the kriging error map obtained from reference dataset shows large errors and where present *B* holes are that have not yet been validated.

For example, on the kriging standard deviation map in Fig. 10, reference holes *A* are in black and *B* holes waiting for validation in red. We will try to validate first *B* holes located in the yellow and orange areas in order to get a maximum standard deviation of 10–12 m.

At each time new holes are validated, it is possible to rebuild the whole geological model automatically.

Cross sections and isopach maps can be derived from the model and are used for verification and for the final validation.

4.3 Automatization

As many drill holes and many formations are to be validated, it is necessary to automatize the work as much as possible.

This is why R scripts are first used for methodology testing.

Once the methodology seems to work well, it is implemented in an in-house software (GDM software, http://gdm.brgm.fr) to be used by geologists who are not specialists in geostatistics.

One of the strengths of this software is that it is coupled with the geological litho-stratigraphic pile and includes many consistency checks. It includes tools for managing the results of cross validation (*set A* with *set A*, *set B* against *set A*), building the model, generating and updating cross section or isopach maps, and

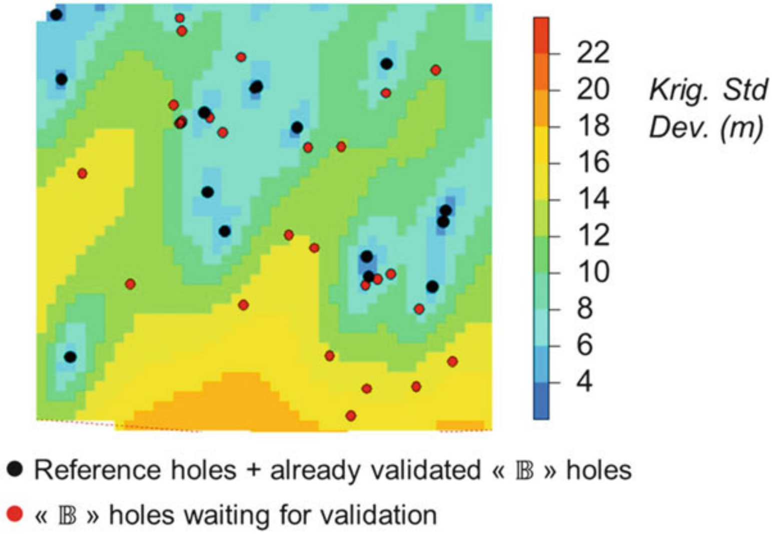


Fig. 10 Use of the kriging standard deviation map to define the next B holes to be validated

managing gaps in case of erosion or formation pinching. However, it does not yet handle variogram automatic fitting nor drift degree identification. This work is still left to the geologist and requires a short training.

This software is able to handle several thousands of drill holes and work on large areas. But in practice, it is preferable to work on a limited area, not for performance reasons, but rather to consider relatively homogeneous areas where the drift degree is constant or where the geological context does not change much.

4.4 Results

Examples of results are given for two areas: the North Aquitanian basin and the Paris basin.

In the North Aquitanian basin (Fig. 11), the geological context is a carbonate ramp, and we looked at Jurassic formations. In this area, 117 reference holes were selected (set A), and 10 formations of the Jurassic were considered.

For the Toarcian, which is one of those ten formations, 60 drill holes were to be validated (set B).

Two thirds of these drill holes could be automatically validated by the procedure, so only one third was left for manual verification.

For control purposes, they were all checked: four were discarded, seven were validated, and nine were erroneous but could be reinterpreted.

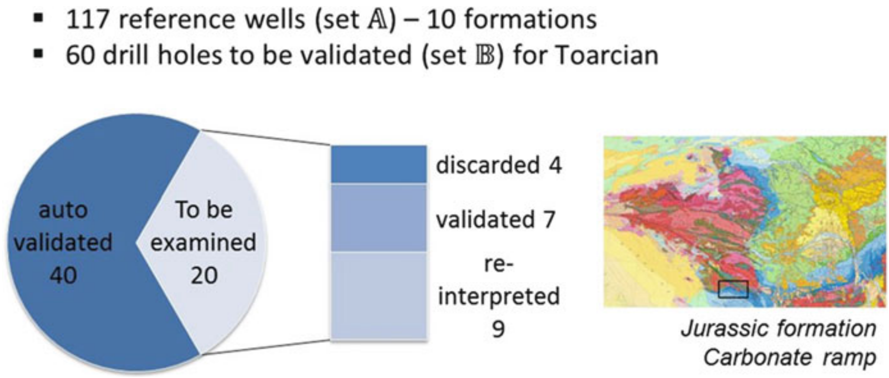


Fig. 11 Result in North Aquitanian basin and location of the study area in France

With the new validated holes, we obtained a reduction of 30 % of the average kriging standard deviation, compared to the situation with only the set *A* holes.

In such simple geological context, drill holes are of good quality and can be easily validated.

The second area where we tested the methodology is in the Paris basin, next to Paris city (Fig. 12). Here we looked more precisely at Eocene formations which deposited in shallow sea or lacustrine context. In this area there are 7000 holes to validate, represented by the black dots in the map. 168 drill holes with gamma ray were interpreted and were used as the reference set (Fig. 4). They are shown in red on the map.

Among the 7000 holes, only 137 intersect the Champigny formation that was selected for the test, and nearly 50 % of holes could be validated automatically. They are shown in blue on the map. These holes are close to reference drill holes because we introduced a low value for the maximum allowable estimation error.

In this case where the geological context is more complex, we can auto-validate only 50 % of holes for the Champigny limestone formation. As there are 19 other formations to validate, the work takes more time, but is facilitated by a semi-automatized workflow.

5 Conclusion

A methodology has been set up in order to enable drill holes validation. It is based on geological concepts mixed with basic geostatistical tools.

This methodology helps finding quickly consistent drill holes and discarding erroneous ones.

As we are considering large datasets and many geological formations, automatic and semiautomatic tools have been developed to save time and ensure repeatability.

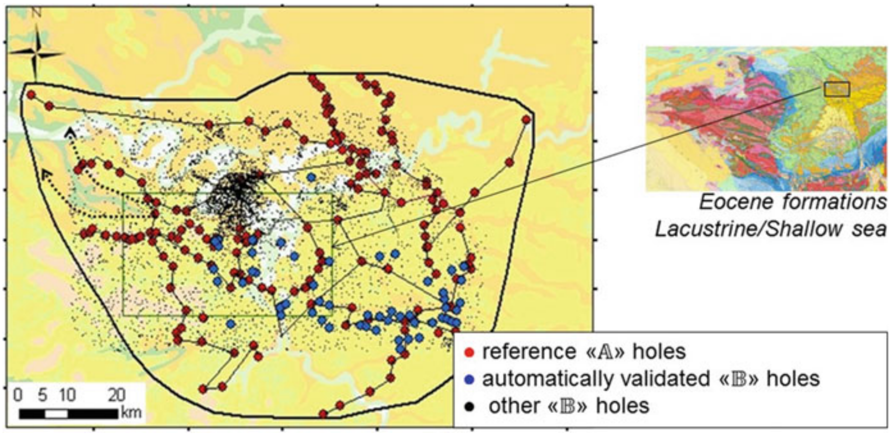


Fig. 12 Result in Paris basin. Case of the Champigny limestone formation

These tools have been made accessible to the geologists in charge of validation and who are not geostatisticians.

Further work to be done concerns (1) improvement in the automation (e.g., automatic looping on formations and automatic database management when adding validated *B* holes to dataset *A*), (2) automatic variogram fitting and drift identification, and (3) finding a way to manage drill holes that have not been automatically validated nor discarded by the present procedure. These holes are kept as "not yet validated" but are not rejected. They are candidate for a further validation, but it would be useful to assign to these drill holes some kind of probability index of being consistent with present knowledge and develop other methods to assess their quality.

References

Abrahamsen P, Benth FE (2001) Kriging with inequality constraints. Math Geol 33(6):719–744

Bourgine B (2015) MultiLayer, Manuel de référence, version 2014. BRGM Report RP64115-FR

Bourgine BLM, Lembezat C, Thierry P, Luquet C, Robelin C (2008) Tools and methods for constructing 3D geological models in the urban environment. The Paris case. In: Ortiz JM, Emery X. GEOSTATS 2008. Santiago, pp 951–960

Freulon X, De Fouquet C (1993) Conditioning a Gaussian model with inequalities. In: Soares A (ed) Geostatistics Troia 92, vol 1. Kluwer Academic, Dordrecht, pp 33–51

Hengl T, Heuvelink G, Rossiter D (2007) About regression-kriging: from equations to case studies. Comput Geosci 33:1301–1315

Homewood P, Mauriaud P, Lafont F (1999) Best practices in sequence stratigraphy for explorationists and reservoir engineers. Bull centre Rech Elf Explor Prod Mem 25:81, pp 50