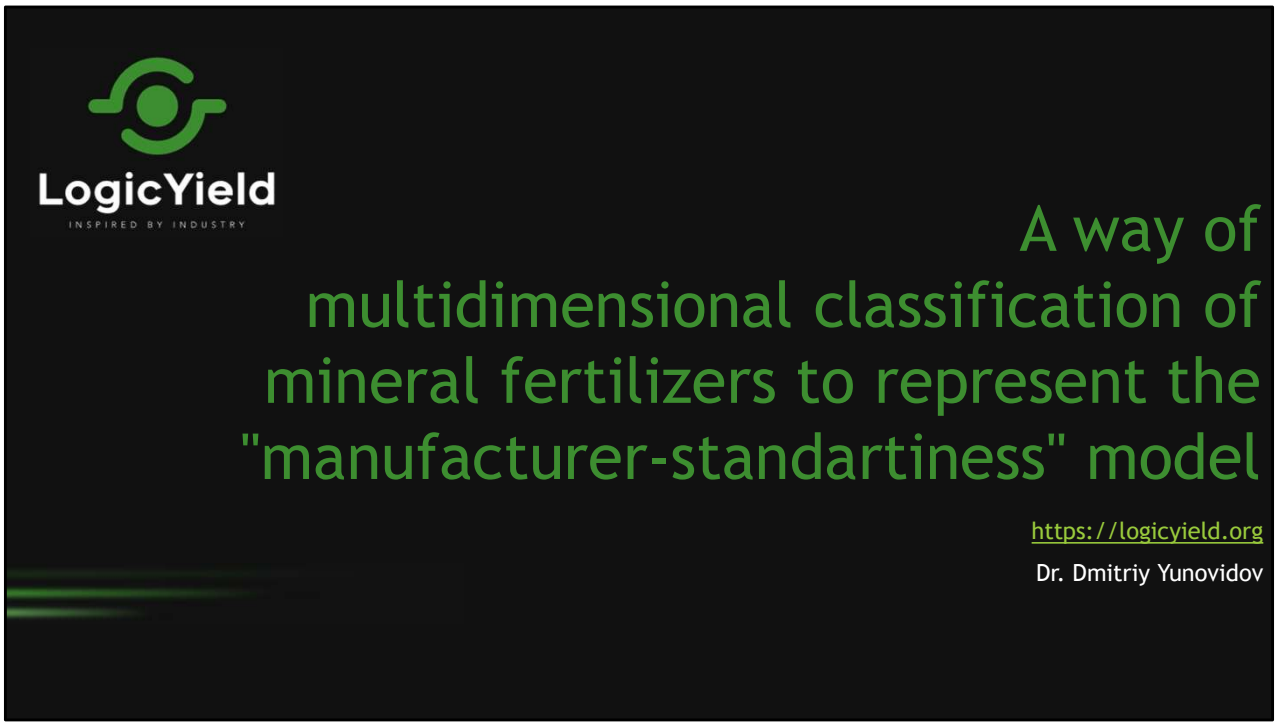



Dear colleagues!

My name is Dmitriy and I am glad to introduce you our research, which is based on Energy dispersive X-ray florescence analyses with principal component analyses and clusterization. Name of the work you can see on the slide.

LogicYield

## Introduction

LLC “LogicYield” studied the mineral fertilizers industry to develop and implement industrial and laboratory robotic quality control systems

- **Objective**: to build a "manufacturer-standartiness" model to identify fertilizers by brand and territorial affiliation.
- **Methods**: energy dispersive X-ray fluorescence analysis (ED XRF), downscaling and k-means clustering.
- **Tasks**:
  - Describe mineral fertilizers
  - Develop technique for calculate “manufacturer-standartiness” model
  - Choose fast and informative method for obtain data
  - Describe application of proposed technique and obtain some result

2

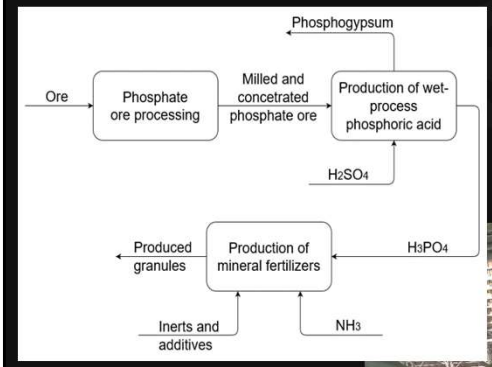
Our company develop devices for quality control in industry laboratory and workshops as well as study big-data approach for such devices based on our own software. Our main field of study is production of mineral fertilizer and you can see here our optical control device for granules size, color and shape for example. And of cause we study Big-Data approach for other methods too.

That is why today I want to discuss with you our scientific research in energy-disperse x-ray fluorescence analyses, which is related to control of brand and producer of fertilizer on the market. As you may know, a lot of counterfeit products and defective mineral fertilizer can be find in the sales markets. Proposed technique of mineral fertilizers multidimensional classification with X-ray fluorescent spectra can help you rapidly define the manufacturer of the unknown sample and draw a conclusion on a single concept of “quality” for the given object. In “quality” we mean set of physical and chemical properties of the object, which are coded in XRF spectra related to the cluster of known samples.

So, you can see here our main objective and methods. Moreover, we will discuss quality of mineral fertilizers, XRF method and some experimental results, which were obtained with proposed technique.

LogicYield

## Mineral fertilizers



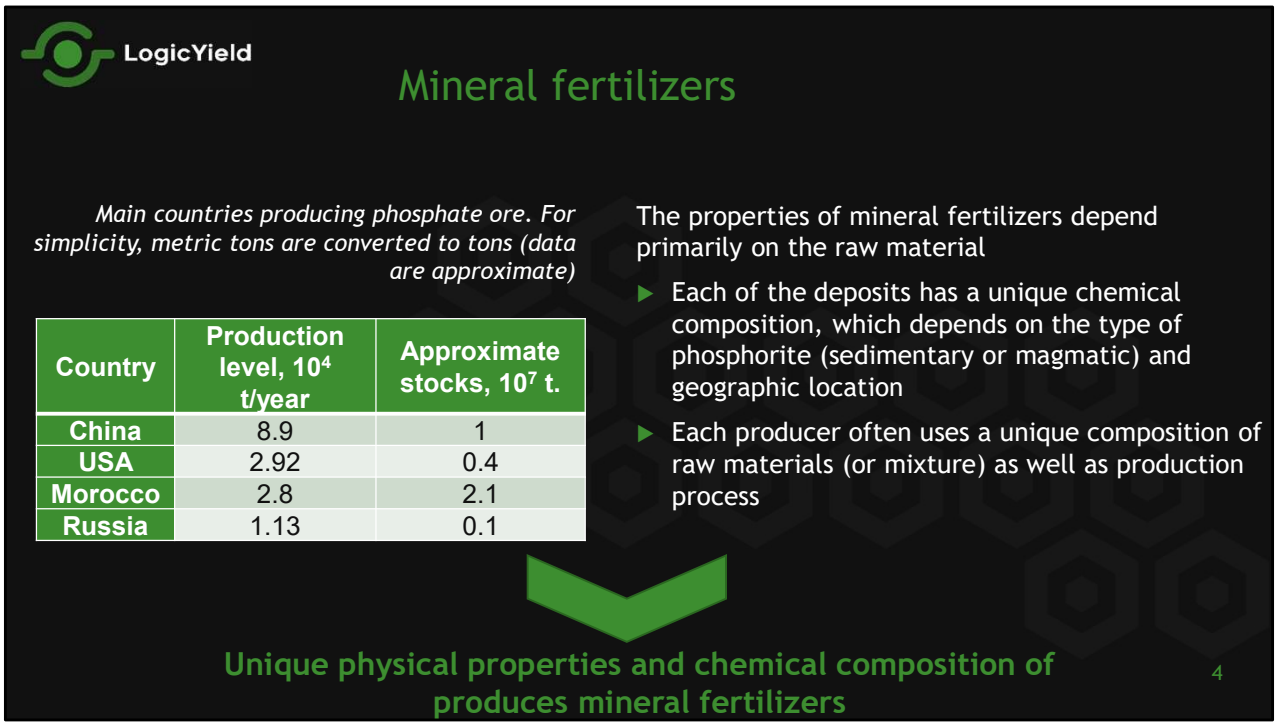

- Mineral fertilizers are a sustainable and growing market for agricultural and ecosystem development (P2O5 consumption > 39 million tons)
- The quality of this market needs to be ensured
- Balanced plant nutrition unique to different types of soils and plants is important
- Depends of production scheme and ore type

3

Modern industrial production is a complex process with many factors, which affect the finish product. No exception and for mineral fertilizer industry either. Market of phosphorus and fertilizers is grow and correlated with Earth population. Moreover we need to ensure the quality and manufacturer of fertilizer brand to avoid counterfeit products and consequences. For example, counterfeit product can contaminate soil with heavy metals (cadmium for example).

You can see here a simple scheme of phosphorus mineral fertilizer production, which shown how finish product depends on ore, sulphur acid, ammonium, different additives and etc. Moreover, each type of soils and each type of plants demand they own nutrient proportion in fertilizer brand. That is why we have a lot of producers and brands on the market.

This is why the quality must include not only the chemical properties, but the whole bunch of physical and chemical properties of selected brand of fertiliser.




According to different sours and type of ore, as well as different type of production, we can conclude, that the properties of finished product are mirrored all these hidden properties. Moreover, if we use data-reach method for analyze granules of fertilizer on the market, it can be possible to predict not only the quality of this product, but its manufacturer and origin technology as well. In other words, we can sole the reverse task.

You can see here some information of main ore deposits, each of them have unique chemical and physical properties.

LogicYield

## Methods of expressing the quality of mineral fertilizers

Quality characteristics of fertilizers:

- nutrient composition (type/brand of fertilizers, e.g. NPK 15-15-15)
- content of microelements (Zn, Mg, etc.) and heavy elements
- solubility, crumbliness, etc.

For efficient product classification we use ED XRF:

- wide range of detected concentration (0.0001 - 100 wt%) and multi-element
- simple sample preparation and fast
- easy automation, etc.

| 1. Samples preparation | 2. Measurement | 3. Calculations |
|---|---|---|
| • 1 - 10 minutes<br>• ~ 80 % errors | • 5 - 300 seconds<br>• ~ 5 % errors | • 1 - 60 seconds<br>• ~ 15 % errors |

5

In our study we use energy-disperse X-ray fluorescence spectrometry, because it is very data reach method. You can see here our laboratory system with optical control of granules. This optical additive give us possibility to analyze colors and shape of pressed granules, which are related to manufacturer as well. But this is theme for other discussion. For now we only use XRF method to record spectra.

Each spectra contain information about elements from Na to U as well as density of sample. Moreover the preparation procedure is fast and methods can itself be easily automated.

And you know from previous slides, that the quality of fertilizer depends on chemical properties, such as nutrient composition or microelement concentration, as well as physical properties, such us density and homogeneous of elements. That is why, XRF spectroscopy is good and data-reach method to analyze the fertilizer from our point of view. We will use it to check proposed model.

LogicYield

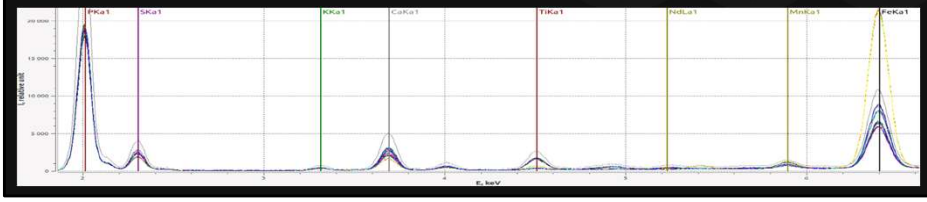
## Multidimensional classification methods

Despite the development of mathematical apparatus and the variety of instrumentation, the XRF method has not found wide application for mineral fertilizer (complex matrix).

We need Data Science to extract information from fertilizer granules:

- Create “object-features” data matrix
- Prepare data in data matrix (standardization, filtration, etc.)
- Use multidimensional statistics to groped data in clusters (PCA + k-means)

No need to calculate classical concentration, just press granules and analyze!





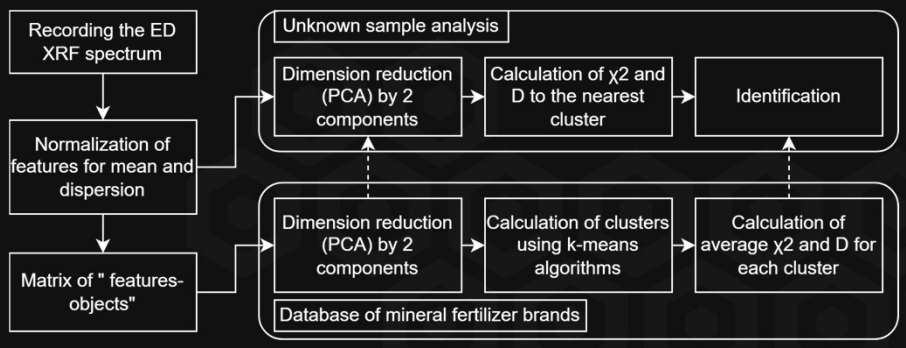
After we select our method to obtain data we need to prepare our math. And this is a hard part. Because of its complicated math and a lot of information is spectra, the XRF method has not found wide application in fertilizer industry yet.

But with multidimensional “big-data” analysis we can extract information from such noise signal as energy disperse spectra. The main idea of such big-data analyses is to create “objects-features” matrix, process it to stabilize the data and finally build clusters to group object according to they manufacturer.

And the best part of it – we do not need to precise calculation or analytical-chemistry regression. We can simply press granules and use whole spectra to calculate our clusters!

LogicYield

## “Manufacturer-standartiness" model




The main idea is to use each XRF spectra channel as feature for PCA with $x^2$ statistics (about 4000 channels in ED-XRF spectra)

7

The main idea of our model present on this slide. You can see base stages to process spectra data and develop database of “knowed” manufacturer with quality products. After we fill our database, we can analyze unknown spectra and match it with database information to find manufacturer and calculate some sort of standartiness of sample or, “quality”. In other words, how this sample related to cluster of good samples. Is it outsider or not.

For such calculation we use Euclidean distance from the center of nearest cluster to unknown spectra in principal component space. Moreover, we use the chi-square statistic to make our data more precise.

LogicYield

1. $\chi^2 \leq \overline{\chi^2}$ и $D \leq \overline{D}$ - the object belongs to the selected cluster and it have the good quality
2. $\chi^2 \leq \overline{\chi^2}$ и $D \geq \overline{D}$ - the object belongs to the selected cluster, however, the PCA is no longer effective (the number of main components needs to be increased)
3. $\chi^2 > \overline{\chi^2}$ и $D > \overline{D}$ - the object does not belong to the selected cluster or it have the bad quality
4. $\chi^2 > \overline{\chi^2}$ и $D < \overline{D}$ - the object belongs to the selected cluster, however, the spectrum is bad (the measurement was incorrect) or some brand properties have changed (e.g. dye has been added to the fertilizer)

The numerical quality value (Δ) is calculated only in first case as

$$\Delta = \overline{\chi^2} - \chi^2 + \overline{D} - D$$

8

For unknow spectra we may have four type of situations, which you can see on the slide. In the first scenario we have perfect sample, which a located inside one of the cluster and chi-square statistic similar to average chi-square for this cluster.

On the other hand, we can have some mismatch with Euclidian distance and chi-square which indicate that we need to add some components to our PCA model (in case if we analyze a lot of similar manufacturer, which is unlikely) or we need to check our XRF spectrometer.

And if both of parameters indicate outsider (case 3 on the slide), we can conclude that we analyze counterfeit product.

Moreover we can calculate our “quality” parameter as single value and use it as approximation for “how far our sample from average product of this brand and manufacturer”.

LogicYield

## Experiment

We studied about 15 samples of each of 4 fertilizers brands according to its manufacturer:

- Met NPK(S) 4-30-15(16) (Volhov branch of JSC «Apatite»);
- Met NPK(S) 0-20-20(5) (Volhov branch of JSC «Apatite»);
- Apa NP(S) 12-40(10) (JSC «Apatite», Cherepovets city);
- Apa NPK 16-16-8; Apa NP 12-52 (АО «Апатит», г. Череповец).

- Analytical mass of granules (200 g) reduce to 50 g and randomly select 3 portion of granules to form a pellet (2 g)
- The granules of each sample were pressed as a "sandwich structure" on a boric acid substrate (280 bar = 11 t/cm2, 30 s)
- Use all ED XRF spectra as features for PCA (at 20 kV, 1000 μA and 100 s in a vacuum)
- Prepare the data base of known samples

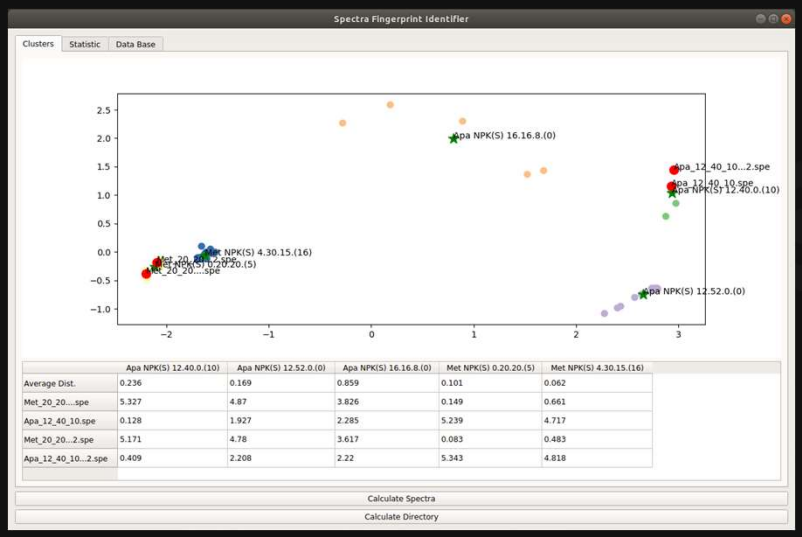
9

Our experiment include creation of small database according to two similar manufacturer and four brand of fertilizers. These two facilities belong to one company and use similar ore to produce fertilizers.

We simply press sample of granule on boric acid and analyze this sandwich-structure in vacuum. After that we prepare four encrypted samples, which are not included in database. You can see example of sample in the picture.

LogicYield

## Results and discussion




| | Apa NPK(S) 12.40.0.(10) | Apa NPK(S) 12.52.0.(0) | Apa NPK(S) 16.16.8.(0) | Met NPK(S) 0.20.20.(5) | Met NPK(S) 4.30.15.(16) |
|---|---|---|---|---|---|
| Average Dist. | 0.236 | 0.169 | 0.859 | 0.101 | 0.062 |
| Met_20_20....spe | 5.327 | 4.87 | 3.826 | 0.149 | 0.661 |
| Apa_12_40_10.spe | 0.128 | 1.927 | 2.285 | 5.239 | 4.717 |
| Met_20_20...2.spe | 5.171 | 4.78 | 3.617 | 0.083 | 0.483 |
| Apa_12_40_10...2.spe | 0.409 | 2.208 | 2.22 | 5.343 | 4.818 |

- We created special program for described calculations (Python, numpy, pandas, sklearn, scipy, matplotlib, PyQt5)
- According to proposed scheme 5 encrypted samples were analyzed (red dots)
- Distance and “quality” parameters to nearest cluster were automatically calculated

10

For process our data and make a conclusion, we created simple program, which can record spectra, create “objects-features” matrix, preprocess data and calculate all statistics we need. You can see our clusters here, as well as unknown samples, which are indicated as red dots.

Moreover, you can see how some of the fertilizers brand are similar to each other and some are not. It is completely normal for such big variety of brands, even if they produced from one type of ore by similar facilities.

According to the obtained data, there is a clear separation of the studied samples in the space of two main components.

LogicYield

| № | Nearest cluster | D | $x^2$ | $\overline{D}$ | $\overline{\chi^2}$ | Δ | Real brand |
|---|---|---|---|---|---|---|---|
| 1 | Apa NP(S) 12-40(10) | 0.128 | 0.164 | 0.236 | 0.304 | 0.248 | Apa NP(S) 12-40(10) |
| 2 | Apa NP(S) 12-40(10) | 0.409 | 0.504 | 0.236 | 0.304 | - | Apa NP(S) 12-40(10) |
| 3 | Met NPK(S) 0-20-20(5) | 0.149 | 0.124 | 0.101 | 0.120 | - | Met NPK(S) 0-20-20(5) |
| 4 | Met NPK(S) 0-20-20(5) | 0.083 | 0.091 | 0.101 | 0.120 | 0.047 | Met NPK(S) 0-20-20(5) |

- According to the results, it was possible to identify each of the samples by brand, but not all of them were of high quality.
- The developed model was realized and the possibility to calculate numerical criterion of the generalised concept of quality was achieved.

11

You can see our result for encrypted samples here. We could clearly identify each of the sample in the space of two main components and calculate value of quality, which we discussed early. Some samples were not so good as expected. It can be because of sulfur, which were added as elemental sulfur and not homogenously dissolve in fertilizer substance.

As you can see, proposed model can rapidly and precisely show manufacturer of unknown sample as well as present numeric value of its quality in terms of “how far our sample from average fertilizer of define manufacture and brand”.

LogicYield

## Conclusion

1. The work describes the approach to express identification of mineral fertilizers based on the “manufacturer-quality” model.
2. The developed model makes it possible to determine the producer of mineral fertilizers and quantify the single concept of "quality" according to the data of ED XRF spectrum (as a set of standard physical and chemical properties).
3. Proposed approach use all of energy channels of the spectrum as separate features (about 4000) and calculate “quality” and identify of the manufacturer based on Euclidean distance with $x^2$ criterion.
4. The developed approach can be widely used in surveyor activities and for the rapid identification of mineral fertilizers on the markets.

12

Finally, you can see our conclusions here. I want to pay attention that we do not propose strict algorithm or mathematical model for multidimensional quality calculation. Instead we propose fast and understandable experimental technique for identification of counterfeit production right in the market in 5 minutes.

Thank you again for you attention and I will gladly answer on any questions.

LogicYield

INSPIRED BY INDUSTRY

# Thank you!

Dr. Dmitriy Yunovidov

+79611121089, Dm.Yunovidov@gmail.com

https://dimyun.space/

13