Afghanistan opium survey report 2022 - Methodology

This chapter covers various methodological aspects regarding survey design and estimation procedure.

1.1 Estimation of area under opium poppy cultivation

Remote sensing methodologies have been used by UNODC since 2002 to monitor the extent of opium poppy cultivation in Afghanistan. Changes in the location of opium poppy cultivation and the increased security difficulties involved in accessing the area of interest require continuous improvements of the survey designs.

A sampling approach is used to cover those provinces where most of the poppy is found, whereas a targeted approach is used in provinces with a low level of opium poppy cultivation. "Targeted approach" means that a certain area of a province is fully covered by satellite imagery. Provinces without indication for opium poppy cultivation are covered by the village survey only.

From 2015, new and better satellite technology allowed for a major change in the study design: the size of the grid cells used for acquiring satellite imagery has been reduced from 10 x 10 km images to 5 x 5 km images. This change affected only provinces where a sampling approach was used; all other provinces were not affected by this change.

In 2022, out of 34 provinces in Afghanistan, 17 were sampled and 11 were targeted. The remaining 6 provinces were considered to be poppy-free based on information from the field.

The same sampling locations have been used since 2019, which ensured high levels of comparability of the annual estimates.

Table 1 Area estimation method, by province, 2022

Region	Targeted approach	Sampling approach
Central	Kabul, Parwan, Logar	Day-Kundi
Eastern	Kapisa, Nuristan	Kunar, Nangarhar, Laghman
Northern	Samangan	Faryab, Jawzjan, Balkh, Sari-Pul,
North-eastern	Baghlan, Takhar, Kunduz	Badakhshan
Southern	Ghazni	
Western	Hirat	Badghis, Farah, Ghor
South-western		Hilmand, Kandahar, Uruzgan, Zabul, Nimroz,

1.1.1 Study design

1.1.1.1 Sampling frame

The sampling frame was established by extracting the area of land potentially available for opium poppy cultivation in 17 provinces. This area was divided into regular 5 km by 5 km grids, which constituted the sampling frame. The final sampling frame, from which the satellite images were randomly selected, consisted of 7,477 cells. In the case of images that cut across provincial boundaries, only the part falling into a particular province was considered to be in that province.

The area available for agriculture in the sampling frame covers irrigated and rain-fed land. The total area in the 17 provinces was 48,810 km^2, which is equivalent to 38% of all potential agricultural land in Afghanistan. Potential land refers to all land available for cultivation and also includes land that is currently fallow.

Cells containing less than 0.25 km^2 of potential agricultural land were excluded from the sampling frame in order to reduce the likelihood of choosing cells with very little arable land. In total, the exclusions represented less than 1% of the total potential agricultural land.

1.1.1.2 Sample size determination

The total number of images to be selected in the sampled provinces was determined in 2015 with the goal to increase accuracy of the estimates and to save cost when compared to previous years.

The accuracy of area estimates depends on the proportion of land covered by satellite imagery and even more so on the number of images than can be acquired. With opium poppy cultivation being concentrated in hot spots and thus unevenly distributed across the agricultural land, information from a large, contiguous piece of land has less value than geographically evenly distributed, smaller pieces information. Costs associated with satellite imagery depends mainly on the total area covered (and not on the number of images). By using 5 x 5 km instead of 10 x 10 km images, at same costs four times the number of images can be acquired. Further details on the sample size determination methodology can be found in O*pium Survey, December 2015*, page 42.

1.1.1.3 Sample size allocation

The available number n of images has been distributed to provinces h according to a so-called power allocation, which uses agricultural area as size measure. For provincial sample size n_h,

$$n_h = n\frac{X_h^q CV_h}{\sum_{h=1}^{H} X_h^q CV_h}$$

where CV_h is the coefficient of variation of area under poppy cultivation in province h and X_h land available for agriculture in province h. This approach ensures that sample size depends on both the variability of poppy and the size of the province measured by agricultural land. After an empirical assessment, the smoothing parameter q, $0 \leq q \leq 1$, *w*as set to 0.2. In addition, a minimum of 20 samples per provinces was set, which took effect in Day-Kundi and Kunar.

Table 2 Sample size and agricultural land and sampling ratio, by province, 2022

Province	Total arable land (km2)	Frame	Effective sample size	Arable land in selected cells	% of arable land represented by selected cells
		# cells	# cells	(km2)	
Badakhshan	3,490	396	53	456	13%
Badghis	6,290	636	50	830	13%
Faryab	7,970	532	86	1426	18%
Jawzjan	3,440	294	39	530	15%
Laghman	263	103	25	61	23%
Ghor	1,615	1144	83	114	7%
Day Kundi	672	406	20	29	4%
Farah	2,820	604	46	480	17%
Hilmand	5,849	696	98	1283	22%
Kandahar	3,472	695	80	763	22%
Kunar	293	124	24	50	17%
Nangarhar	1051	181	26	334	32%
Nimroz	1,083	213	36	342	32%
Balkh	4577	256	40	865	19%
Saripul	3557	379	56	600	17%
Uruzgan	903	277	30	89	10%
Zabul	1,465	541	29	150	10%
Total	**48,810**	**7,477**	**821**	**8,402**	**17%**

1.1.1.4 Sample design

In 2015, MCN/UNODC undertook an extensive simulation study which compared various sampling designs and estimation methods in order to determine the best (most accurate with a given number of samples) design for a certain situation.

Case studies were undertaken for Hilmand and Kandahar province. The sampling designs considered have been used in the past by MCN/UNODC:

- simple random sampling,
- probability proportional to size sampling (PPS), using agricultural area as a size measure,
- stratified random sampling using compact geo-strata of equal size as strata,
- systematic random sampling.

Two estimation methods have been compared: a ratio estimator using agricultural area as auxiliary variable and the Horvitz-Thompson estimator.

The study concluded that for the two cases considered

- PPS performed best, and
- The ratio estimator is to be preferred for simple random sampling, systematic random sampling, and stratified random sampling. For PPS, it does not yield any improvements in accuracy.

The PPS builds on the correlation between the size measure and the variable of interest. In provinces where poppy and agricultural land are highly correlated, PPS is expected to perform best. In provinces, however, where poppy and agricultural land are only weakly correlated, PPS does not bring any advantages and might reduce accuracy.

Therefore, in Badghis, Balkh, Farah, Faryab, Hilmand, Kandahar, Nimroz and Zabul, PPS was applied. In the remaining provinces, systematic random sampling was used, a sampling design that ensures an even geographical distribution of samples (see the “Opium poppy 2015 – Cultivation and production” for more details).

In more detail, in a PPS design without replacement a unit has a probability to be selected in the first draw of

$$pi = \frac{x_i}{\sum_{i=1}^{N} x_i}$$

where *x* is the size variable (agricultural land) in unit *i*, and *N* is the number of units that can be selected. The subsequent units have slightly modified inclusion probabilities. For drawing the samples and for calculating the inclusion probabilities the statistical software *R* (package *sampling*) was used.

Since agricultural area tends to be concentrated in one or more clusters in a province, PPS sampling without further stratification would lead to a concentration of samples in a few spots and possibly do not cover every district. Therefore, in all PPS provinces, the sample was stratified by district.

In the remaining provinces, a one-stage systematic random sampling approach was employed in which a sampling rule was applied that ensured good geographic coverage. Starting from a randomly chosen cell, every *k*th element from then onwards was chosen, where *k* is determined by the number of cells in the frame and the desired sample size (the actual sample size might differ slightly).

In *Nangarhar* province, the districts Dara-e-Nur, Kuzkunar, Kama, Behsud, Jalalabad and partially Surkhrod were excluded from the frame.

1.1.2 Area estimation in sampled provinces

The estimation of the extent of opium poppy cultivation is a ratio estimate[1] for each of the provinces, using potential agricultural land as an auxiliary variable. The national estimate was obtained by adding up the provincial estimates in what is known as a separate ratio estimate.

In provinces where systematic random sampling was applied, the area of opium poppy cultivation, Yk, within province *k,* is estimated as:

$$Y_k = X \frac{\sum_{i=1}^{n_k} y_i}{\sum_{i=1}^{n_k} x_i}$$

where n_k is the number of satellite image locations within the province; y_i is the area of poppy cultivation in image *i;* x_i is the area of land potentially available for poppy cultivation in image *I*, and *X* is the total potential land available for poppy cultivation in province *k.*

In PPS provinces, where units are selected with unequal inclusion probability, a slightly different ratio estimate was used that incorporates the inclusion probability (Horvitz-Thompson estimator).

1.1.2.1 Uncertainty

In the PPS provinces the confidence intervals were calculated following statistical practice.[2]

In all remaining provinces no unbiased estimator for the variance was available; confidence intervals were approximated by assuming simple random sampling. The confidence intervals therefore slightly overestimate the uncertainty of the estimates.

[1] The ratio estimator did not outperform the Horvitz Thompson estimator in the PPS provinces. The ratio estimator was applied in all provinces for reasons of consistency and to account for possible updates of the agricultural area in future years.

[2] See, e.g. Cochran, W. G., Sampling techniques, John Wiley & Sons (2007).

Table 3 Area estimates of sample provinces with 95% confidence interval, 2022 (Hectares)

Province	Point estimate (Hectares)	Lower bound (Hectares)	Upper bound (Hectares)
Badakhshan	4,305	1,661	6,950
Badghis	14,110	6,149	22,071
Balkh	4,542	1,896	10,813
Day kundi	837	350	2,600
Farah	15,829	8,950	22,707
Faryab	6,929	4,522	9,335
Ghor	1,784	857	2,710
Hilmand	122,045	112,877	131,213
Jawzjan	1,359	383	2,335
Kandahar	29,229	23,135	35,323
Kunar	822	423	1,452
Laghman	1,102	446	2,188
Nangarhar	5,241	4,219	6,263
Nimroz	2,429	1,601	3,257
Saripul	3,454	1,314	5,594
Uruzgan	14,403	8,208	20,599
Zabul	1,531	595	4,166

To express the uncertainty associated with the national area estimation, which includes the provinces covered by the targeted approach and the sample provinces, but excludes provinces with an estimate of less than 100 hectares (which are considered "poppy-free" and not counted), a range was calculated by adding the poppy area figures of the target provinces to the upper and lower limits of the 95% confidence interval at the national level.

1.1.3 Area estimation in target provinces

The consensus view of those working in Afghanistan was that the surveillance system developed in the provinces can identify sites where poppy was grown, with further inputs being obtained from the survey of village headmen. Fieldworkers visited potential poppy-growing sites to confirm the situation and provided GPS references for the sites. If geographical clusters of sites were identified, targeted satellite images were obtained to measure the areas involved. The total poppy area of a target province is equal to the poppy area measured on the imagery without any further calculation. For a list of provinces for which the target approach was used see respective table.

In provinces where satellite images were targeted, the estimated area under opium poppy cultivation is not affected by sampling errors, although they may be affected by the omission of areas with very little cultivation. Area estimates of target provinces should therefore be considered as a minimum estimate.

1.1.4 District level estimation

District level results are indicative only. For district level estimation all cells are used which have the majority of agricultural area in that district. That means that in certain cases, agricultural area and poppy cultivation is accounted for in a neighbouring district and not within the district where cultivation occurred. This is, however, in most cases set off by those cells, where the contrary is the case.

1.1.5 Accuracy assessment

Due to the difficult security situation in many parts of Afghanistan, which prevented surveyors from carrying GPS and mapping equipment, an insufficient number of ground segments could be visited in order to conduct a systematic accuracy assessment.

1.1.6 Estimation of the net cultivation area

In years where eradication took place and was verified, the area figure presented is the net harvestable opium poppy cultivation area. The effect of poppy eradication activities was taken into account based on data from the eradication verification survey, which provides exact GPS coordinates of all eradicated fields supplemented with additional information. The gross cultivation areas would be the net cultivation plus eradication.

In provinces where the poppy area is estimated with a sampling approach, the first step is to calculate the gross poppy cultivation area. The total area eradicated in those provinces is then deducted from the mid-point estimate of the provincial cultivation estimate to obtain the net cultivation area. If eradication activities were carried out after the date of the image acquisition, no adjustment is necessary as the poppy present in the image reflects the gross poppy area. If eradication activities were carried out in a sample block before the date of the image acquisition, the area interpreted as poppy would not reflect the gross area. Therefore, the eradicated fields are added to the interpreted fields. The adjusted poppy area figure for the block is then used for the provincial estimate.

In provinces where the poppy area is estimated with a targeted approach (census), eradication activities that happened before the date of the image acquisition are already reflected, as these fields no longer appear as poppy in the image. Fields that were eradicated after the date of the images acquisition are simply deleted.

1.2 Satellite image interpretation

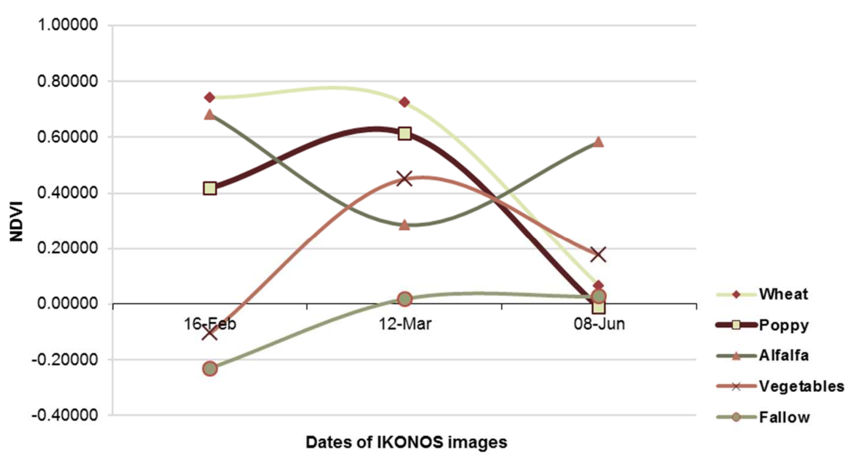
1.2.1 Acquisition of satellite images

The acquisition of satellite images at the appropriate growth stage of the opium poppy is key to the successful identification of opium poppy fields on satellite images. Satellite data is collected at two stages: the pre-harvest (flowering) stage and the post-harvest (post-lancing) stage. In recent years, detailed information on the crop growth cycle of each district has been collected in the form of a phenological chart, which is useful for deciding on appropriate dates for satellite data acquisition. First-dated images of the Southern, Eastern and Western regions are collected during March and April due to the early cultivation and maturity of crops in those regions. The crop growth cycle begins later as one goes northward. Images of the North and North-eastern region are acquired during May, June and July. Second-dated satellite images are collected approximately two months after the first images are collected.

The normal time window for satellite data acquisition is one month, depending on the scheduled passing of satellite and weather conditions. The time window for first-dated image acquisition begins at the full flowering stage and continues through the capsule stage. Second-dated image acquisition begins towards the end of the lancing stage and continues until the opium poppy fields are ploughed. Images acquired in the middle of the prescribed time window facilitate optimum discrimination between opium poppy and other crops.

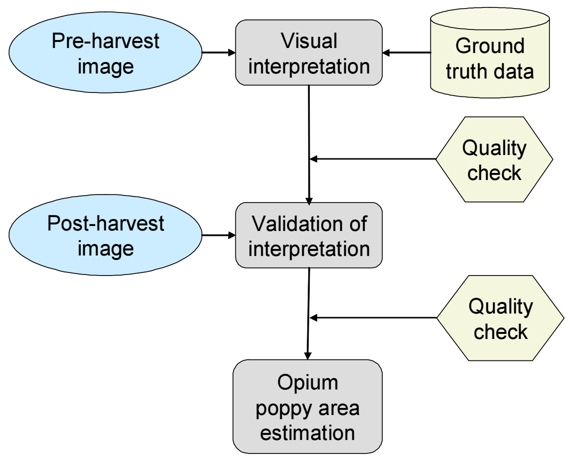
The figure below illustrates the spectral characteristics (expressed in a Normalized Difference Vegetation Index; NDVI) of opium poppy and other crops between February and June. Wheat and opium poppy have the same growth cycle between March and June, as illustrated. The spectral differences between those two crops are more pronounced in February, which marks the beginning of the capsule stage of the crop in this example. Poppy fields are ploughed immediately after the harvest, whereas wheat fields are not. That is why two-dated images (pre-harvest and post-harvest) are collected for the same location.

Figure 1 Spectral reflectance of opium poppy and other crops



The figure above illustrates the growth cycles of opium poppy, wheat and clover from February to June, with the help of ground photographs. Note that maximum visual discrimination between opium poppy and other crops is possible during the flowering/capsule stage and after capsule lancing. The different phenological stages described above are shown in the figure on the previous page (field photographs of opium poppy, wheat and clover on different dates).

Figure 2 Image classification methodology for estimating opium poppy cultivation area



1.2.2 Interpretation of opium poppy cultivation from satellite images

First-dated images were acquired during the flowering or capsule stage and second-dated images were acquired after the opium harvest. For example, wheat appears mostly in bright red on the first date image in false colour composite (full coverage with vegetation appears in red; bare soil in grey/green), while opium poppy fields are shown in tones of pink. Although there can be some confusion between opium poppy and wheat in the first-dated images, the acquisition of second-dated images makes it possible to distinguish opium poppy from other crops, because the opium poppy crop has been harvested and the fields appear in grey/green.

Visual interpretation was used to delineate opium poppy fields by interpreting PLEIADES images covering a 5 km by 5 km area. Ortho-rectified PLEIADES images of 0.5 m resolution (PAN-sharpened) were used for this purpose. Opium poppy was initially identified using first-dated high-resolution images.

Ground truth information collected in the form of segment maps and GPS points was also useful in developing identification keys for opium poppy fields. The interpretation based on first-dated images was improved using patterns of observation in second-dated images. Poppy field boundaries were delineated by an on-screen digitization method.

1.2.2.1 Band combination for opium poppy identification

Two kinds of band combination were used to detect opium poppy. True-colour combination (blue, green, red) was used in areas where land use is dominated by opium (for example, Hilmand and Kandahar) and in cases where images were obtained during the flowering and lancing stages of opium poppy. False-colour combination (infra-red, red, green) was used in almost all cases. Analysts used both combinations simultaneously to optimize discrimination between opium poppy and other crops.

Some of the images could not be acquired at the appropriate time due to weather conditions and/or the time at which the satellite passed. The delayed acquisition of images makes it difficult to detect opium poppy, since fields may be at the senescence stage due to the lancing of capsules and can therefore be confused with fallow fields. In such cases, second-dated images are often useful in confirming opium poppy fields, since harvest patterns are different for wheat and opium poppy.

1.2.2.2 Advantage of two-dated images

In provinces and areas where interpretation was challenging, second-date imagery was obtained. The second-dated images were useful to distinguish poppy from barley, wheat and grapes in certain provinces. The second-dated (post-harvest) images were useful in confirming whether the opium poppy on the first-dated images had been correctly identified. Image acquisition at two different times (pre- and post-harvest) has been proven to be essential in such cases.

1.2.2.3 Quality control

A quality control mechanism was applied to the image interpretation process, with each analyst's work being checked by two other experts. Both first-dated and second-dated images were cross-checked.

All fields determined as likely to be under opium poppy cultivation (potential opium poppy fields) were delineated on the basis of the interpretation of first-dated satellite imagery. In some cases a second-dated image was acquired for the purpose of confirmation. The corrections involved a few commissions and omissions.

1.3 Opium yield and production

In 2022, due to the situation on the ground, it was not possible to collect opium yield data from field measurements. In the absence of the field survey, UNODC applied the methodology developed in 2020 for estimating opium poppy yields using satellite imagery.

In this approach, a representative sample of opium poppy fields was visually ranked according to the quality of crops on the field. Each value of the rank (low, medium, and high quality) corresponded to a specific range of yields, which was estimated from previous years' yield data per region. For example, if previous years' data indicated that poppy fields visually ranked as low quality had yield values of some 10.6 kg/ha, then all poppy fields ranked low quality in 2022 were assigned this yield value. The procedure was repeated for all ranks, stratified by region, resulting in an average yield, weighted by quality of the fields.

The methodology from 2022 was re-visited by evaluating newly collected data from 2021 and adjusting the 2020 method based on new findings from yield data analysis.

In greater detail, the methodology consisted of the following steps:

1) UNODC visually ranked a nationally representative sample of opium poppy fields (from 2015 to 2019, and 2021) in low, medium, and high quality with based on high-resolution imagery.
2) These quality levels where matched with yield values obtained from previous field measurements from the same opium poppy fields and year.
3) Based on detailed statistical testing (Kruskal-Wallis Rank Sum Test), data from the years 2018 (heavy drought) and 2021 (very high yields) were excluded as outliers.
4) The testing further revealed that for rank 1, a single estimate can be applied across all regions, but for ranks 2 and 3, the southern and western regions behaved statistically different and where therefore treated separately.

The yield values corresponding to each quality level and region are below:

	n. obs.	mean	st.error	Bootstrap (1000 samples) 95% confidence interval	
				lower	**upper**
Rank 1 All Regions	37	**10.61**	0.90	8.86	12.31
Rank 2 South	39	**24.58**	1.88	21.03	28.30
Rank 2+3 West	53	**21.65**	1.47	18.89	24.44
Rank 2 Other Regions	94	**34.28**	1.32	31.73	36.95
Rank 3 South	37	**40.62**	2.13	36.37	44.97
Rank 3 Other Regions	132	**47.70**	0.98	45.78	49.72

5) These yield values were used for estimating yields in 2022. First, UNODC visually ranked a nationally representative sample of opium poppy fields in 2022 which resulted in a distribution of poppy field qualities by rank. The corresponding yield values were assigned to each quality level, yielding a weighted average yield per combi-region, which then translated into a national average yield.