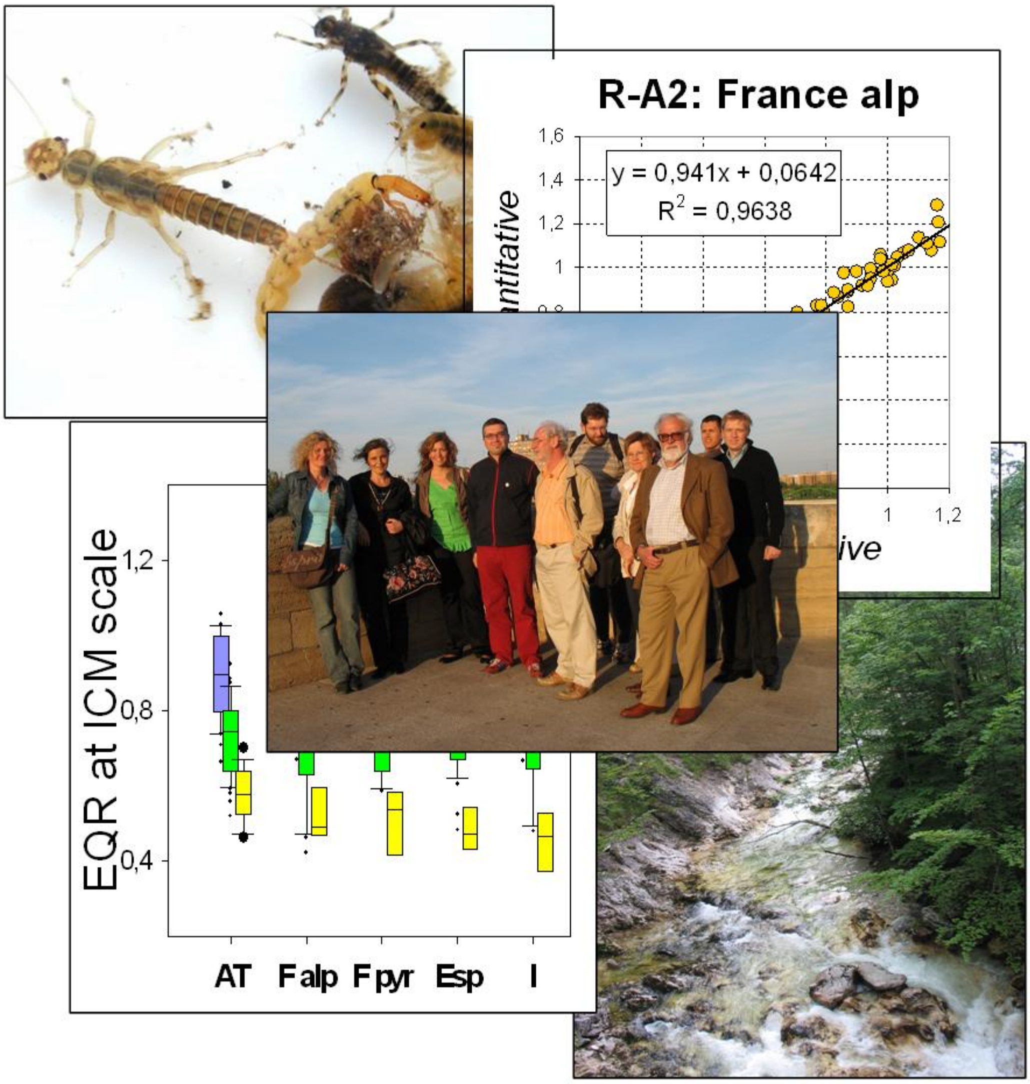
ANNEX C

Alpine GIG

Intercalibration of the boundary values for the macrozoobenthos



Technical aspects of the comparison of the boundary values by using the ICMi – method and Final results

Franz H. Wagner / Austria

ALPINE GIG-MEMBERS

Gisela Ofenböck/AT
Franz Wagner/AT
Jean-Gabriel Wasson/FR
Nicolas Mengin/FR
Ilona Schlösser/DE
Folker Fischer/DE
Paolo Negri/IT
Maurizio Siligardi/IT
Gorazd Urbanic/SI
Bernarda Rotar/SI
Miriam Pardos/ES
Concha Duran/ES

CONTENT

1. Summary

The Alpine GIG used the *Intercalibration Common Metrics index* (ICMi) for the Intercalibration of the boundary values for the HIGH/GOOD and GOOD/MODERATE ecological status classes (macroinvertebrates). The ICMi approach is described in detail in the paper - a simple index which can be calculated with the data of all member states, serving as a comparison tool.

Data limitations, natural variability of aquatic ecosystems and the simplification principles of the ICMi are sources of variability to the results of the intercalibration. Thus it is suggested to use an „*acceptable variation width*“ of the results rather than a single boundary value. The proposal is to use the median boundary values of all member states +/- ¼ of the median good status class range of all member states.

The result of the intercalibration for both alpine stream types seems to be in an acceptable range, especially for the type R-A2. In the type R-A1 Italy is not within the proposed range for the boundary values.

2. Methodology of the Intercalibration in the Alpine GIG

2.1. Short description of the approach

Within the two-year timetable of the intercalibration exercise it is impossible to evolve, test and implement in all member states a common method, or even a common classification based on existing datasets. Thus intercalibration has to deal with existing national methods. Due to the number of MS and variety of methods, it is impossible to use only bilateral comparisons.

No single biological metric can reflect all features required by the normative definitions, i.e. taxonomic composition and abundance, disturbance sensitive taxa and presence/absence of major taxonomic groups, diversity).

As a consequence, a multimetric approach with qualitative and quantitative data was used to take into account these various criteria.

The basic principle of this intercalibration procedure is that all the participating member states submit macrozoobenthos data sets covering as many status classes as possible. The national EQR values are then correlated with EQRs from a simplified method – a multimetric index (ICMi: Intercalibration Common Metric Index) that can be calculated with the data of all countries. This ICMi serves as a comparison tool, like a ruler with a common unit. By using the correlation the national boundary values are then transformed from the national EQR scale into ICMi – EQR values and can easily be compared (Figure 1).

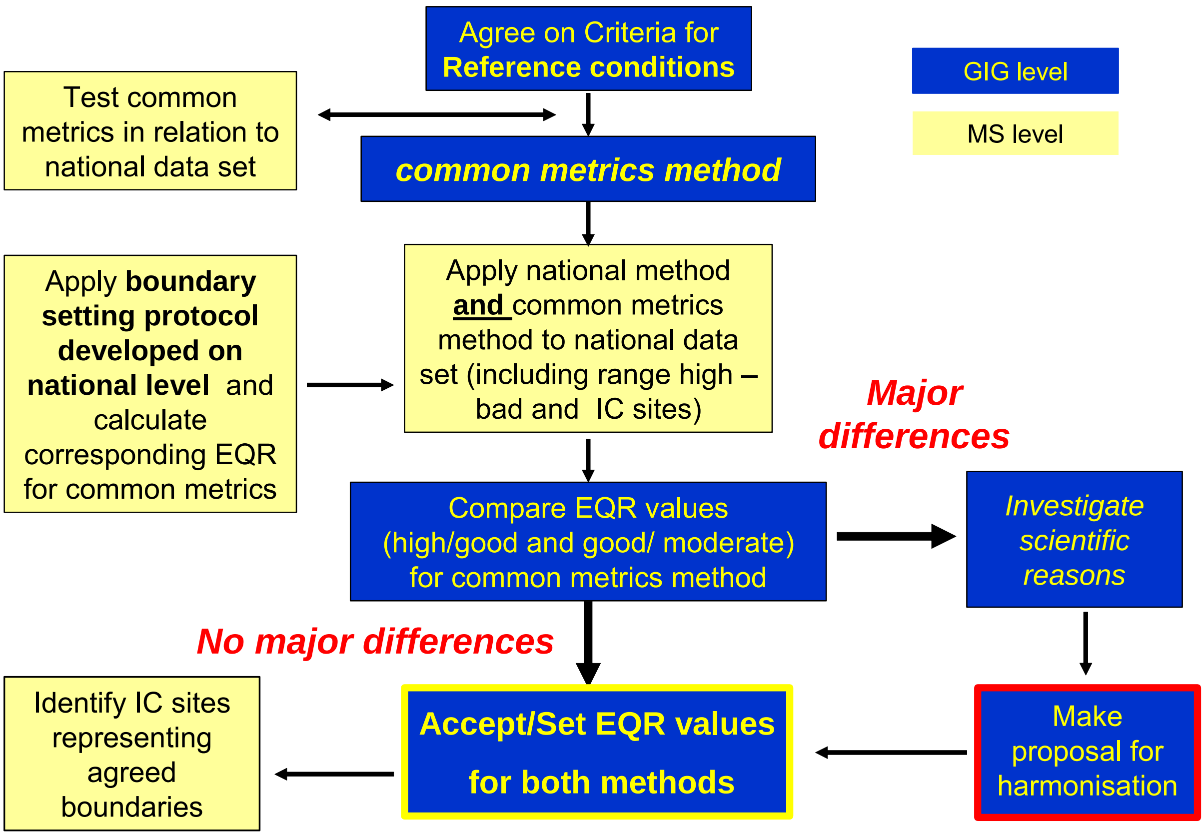


Figure 1: Intercalibration process in the Alpine GIG, GIG Meeting 2004-06-29.

2.2. Typology: IC types and conformity with nominated sites

Two types were selected for the Intercalibration in the Alpine Geographical Intercalibration Group (Table 1) and six countries are participating in the intercalibration process of the Alpine GIG (Table 2).

Table 1: Intercalibration types for the Alpine GIG.

Type	***River characterisation***	***Catchment area (of stretch)***	***Altitude & geomorphology***	***Alkalinity***	***Flow regime***
R-A1	*Small to medium, high altitude calcareous*	10-1000 km^2	800-2500 m (catchment), boulders/cobble	high (but not extremely high) alkalinity	
R-A2	*Small to medium, high altitude, siliceous*	10-1000 km^2	500-1000m (max. altitude of catchment 3000m, mean 1500m), boulders	Non-calcareous (granite, metamorphic). medium to low alkalinity	nival-glacial flow regime

Table 2: Member states participating in the Alpine GIG.

Type	***River characterisation***	***DE***	***AT***	***FR***	***IT****	***ES***	***SI***
R-A1	Pre-alpine - Small to medium, high altitude calcareous	X	X	X	X		X
R-A2	Alpine - Small to medium, high altitude, siliceous		X	X	X	X	

The participating member states nominated sites for each intercalibration type, the sites were screened for typological conformity by collecting descriptive data (Table 3)

Table 3: Descriptive data for the nominated sites of both types

Type R.A1		**FR**	**AT**	**DE**	**SI**
Number of sites		25	55	36	25
Altitude	*min*	160	500	410	180
	max	1002	1140	869	940
	median	395	632	610	555
	mean	478	694	622	580
Catchment area	*min*	54	106	6,8	12
	max	899	909	150,8	94
	median	299	210	30,0	31
	mean	355	292	44,8	32

Table 3: *continued*

Type R-A2		FR-alp	FR-Pyr	AT	ES
Number of sites		17	8	123	35
Altitude	*min*	249	18	380	588
	max	1927	1219	1516	1733
	median	608	374	837	937
	mean	906	502	836	991
Catchment area	*min*	37	12	5	10,4
	max	1047	792	1156	954,6
	median	608	171	62	191
	mean	524	281	169	249

2.3. Reference conditions: site selection and the derivation of reference values

The REFCOND guidance was used as a starting point, where reference conditions are defined as "… a state in the present or in the past corresponding to very low pressure, without the effects of major industrialisation, urbanisation and intensification of agriculture, and with only very minor modification of physico-chemistry, hydromorpology and biology" (REFCOND guidance). The guidance document provides pressure-specific criteria for such conditions. The GIG has reviewed these criteria, and agreed on a common interpretation (see Annex B), A description of the national criteria for reference conditions is included in the Annex A.

Individual sites qualifying as potential reference sites according to the pressure criteria were further screened by applying the pressure criteria specified above, using expert judgement. An additional requirement was that sufficient data should be present to calculate both the national metric(s) and the intercalibration metrics.
The number of suitable reference sites for each common intercalibration type and each country is given in Table 4:

Table 4: Number of reference sites for all intercalibration types and countries. For some sites several samples were available.

Country	Number of reference sites	
	R-A1	***R-A2***
Austria	7	7
France alpine	4	21
France pyrenean	-	16
Germany	2	-
Italy	14	14
Slovenia	5	-
Spain	-	3 (12 samples)

For every metric used in the ICMi the reference value was calculated as median of the metric values from the reference sites).
The reference sites selected by the member states for the ICMi comparison were not in all cases identical with the reference sites for the national methods. Additionally, the national calculations of reference values differed in some cases from the approach used for the intercalibration. This adds difficulties to the data interpretation of the results (see 4.4.)

2.4. Relationship between pressure and the selected metrics

Unlike lakes, in rivers no abrupt changes do usually occur when the Good/Moderate boundary is exceeded. Especially when a combination of pressures is acting on a river site (i.e. the most frequently observed situation), a continuous alteration is found along quality gradients. In general, in rivers the important changes in the benthic community functional structure occur at a point considerably lower than the G/M boundaries proposed by most member states. There is no break point or clear threshold in the pressure- impact relationship of relevant metrics in the range of high and good status sites.

The functional structure of the fauna (e.g. in terms of habitat-related or trophic characteristics) is not drastically altered.

The WFD normative definitions for river macroinvertebrates concerning the good and moderate status are specified in Annex V, 1.2.1:

Good status	**Moderate status**
There are slight changes in the composition and abundance of invertebrate taxa from the type-specific communities *The ratio of disturbance sensitive taxa to insensitive taxa shows slight alteration from type-specific levels.* *The level of diversity of invertebrate taxa shows slight signs of alteration from type-specific levels.*	*The composition and abundance of invertebrate taxa differ moderately from the type-specific communities.* *Major taxonomic groups of the type-specific community are absent.* *The ratio of disturbance sensitive taxa to insensitive taxa, and the level of diversity, are substantially lower than the type-specific level and significantly lower than for good status.*

The key terms ***taxonomic composition*** *and* ***abundance***, ***disturbance sensitive taxa***, ***diversity***, and ***major taxonomic groups*** are interpreted as follows:

- ***Taxonomic composition and abundance***: the terms “taxon/taxa” do not refer to a particular taxonomic level; it is understood as any of the various levels commonly used for faunistic evaluation, i.e. the species, the genus or the family for the most common faunistic groups. Different levels can be used for some groups (e.g. Oligochaeta). “Taxonomic composition and abundance” is understood as “qualitative and quantitative data at the taxa level must be used to describe the community”

- ***Disturbance sensitive taxa***: this is understood as “taxa responding to different kinds of stressors”, i.e. organics and nutrients, hydro-morphological alteration, toxic pollution and acidification. Many taxa or metrics respond to a combination of various stressors viewed as “general degradation”

The Alpine Intercalibration common metrics (ICM_{alpine}) include some special insect orders or families, known as “sensitive” against general degradation (based on empirical data over decades). They respond to different kinds of stressors with a decrease of species numbers and individuals, while more insensitive taxa persist or increase in species and individual number.

- ***Diversity*****:** this can be understood in (at least) two ways:

  - Biodiversity, or taxa richness, expressed by the number of different taxa.
  - Ecological diversity, i.e. some formula combining the number and relative abundance of taxa such as the Shannon index, the Pielou index (or equitability component of the Shannon index). Other more simple indices use only the total number of taxa and the total number of individuals (Margalef, Simpson, etc.).

- ***Major Taxonomic Groups*** **(MTG**) are defined as:

  - having a high probability of occurrence in the type specific community;
  - having a functional significance in the community, (i.e. relatively abundant in proportion of the invertebrate fauna).

    In this sense, groups that are occasionally present, or represent a small percentage of the type specific fauna cannot be considered as major groups.

The evaluation of GIG datasets show, that a significant decline in the total number of orders can only be observed for poor and bad status class. The total number of orders can therefore not be directly used for setting the good/moderate boundaries.

EPT-Taxa (Ephemeroptera, Plecoptera and Trichoptera) can be seen as the most sensitive taxonomic groups in aquatic ecosystems. There is a strong correlation between the number of EPT-taxa and anthropogenic alteration. As the decrease in the number of EPT is used in the Alpine ICM it can be concluded, that the absence of major taxonomic groups for the moderate status classification is taken into account and WFD requirements are fulfilled.

A total disappearance of E, P and T-Taxa can only be found below good/moderate class boundary. The disappearance of EPT can therefore not be used for defining good/moderate class boundary.

A more detailed analysis by the different member states concerning the relationship between pressure and the selected metrics are included in Annex A.

Interpretation of the normative definitions within the GIG comes down to

- agreeing how to quantify taxonomic composition, abundance, disturbance sensitive taxa, and diversity, and major taxonomic groups. This has been done by defining an "**Intercalibration Common Metric index" (ICMi**), as described above.
- agreeing on what constitutes a slight and a moderate deviation from reference conditions. Because the normative definitions do not give any clarification of the meaning of 'slight' and 'moderate', and the lack of obvious break points or thresholds, interpretation of 'slight' and 'moderate' is rather arbitrary. The approach followed in the intercalibration process has been to compare the results of each Member State's method to a common set of Water Framework Directive compliant metric, combined in an Intercalibration Common Metric index (ICMi). Where significant differences occurred between Member State's class boundaries, explanations of these differences were required and adjustments were made where necessary

2.5. Selection of metrics for the ICMi

The metrics combined in the multimetric index ICMi must
- reflect the main features
- respond to various stressors
- allow a quantitative evaluation of a "slight deviation" and 'moderate deviation"

Based on previous experience with similar sets of metrics (e.g. AQEM Consortium, 2002; Buffagni et al., 2004a; Hering et al., 2004; Pinto et al., 2004) a number of metrics was initially proposed (Buffagni & Erba, 2004) and tested for their suitability to be used for the procedure of intercalibration in the Central Baltic GIG.

These Intercalibration Common Metrics, or ICMs, (Buffagni & Erba, 2004; Buffagni et al., 2005) can be clustered in two groups:

- ***qualitative metrics***, only using qualitative information
- ***quantitative metrics***, based on abundance estimates.

After discussions and testing in the Alpine GIG the ICMi of the Central Baltic GIG was slightly modified. Due to the unavailability of data on abundance in some countries (IT, ES), quantitative metrics were excluded from the calculation of the ICMi. But as quantitative metrics are used in some of the national methods a second ICMi including quantitative data was calculated.

The taxonomical resolution for the ICMi is the ***family level*** – as this is the level for which all participating member states are having data available. Most member states are using a more detailed determination of the invertebrates at the species level for their national assessment methods.

2.5.1 ICMi alpine – qualitative data

The Alpine ICM index (ICMi) value is calculated by the averaging the EQR – values of the metrics indicated in Table 5.

Table 5: The metrics used in the ICMi qualitative and their indicative power.

KEY METRICS	*Indivative for:*
▪ total number of taxa ▪ number of EPT taxa ▪ number of selected (sensitive) taxa	➢ *taxa richness* ➢ *general degradation* ➢ *habitat degradation*
▪ ASPTIberian - 2	➢ *Organic pollution*

The invertebrate families listed in Table 6 were considered as being sensitive to general degradation in alpine streams - mostly in relation to hydro-morphological pressures. The selection was made by national experts of all member states participating in the Alpine GIG.

Table 6: Selected sensitive taxa –used as a common metric of the ICMi.

Selected sensitive taxa		
Blephariceridae	*Perlodidae*	*Cordulegastridae*
Empididae	*Taeniopterygidae*	*Elmidae*
Ephemeridae	*Beraeidae*	*Hydraenidae*
Heptageniidae	*Brachycentridae*	*Helodidae*
Leptophlebiidae	*Glossosomatidae*	*Astacidae*
Capniidae	*Goeridae*	*Planaridae*
Chloroperlidae	*Lepidostomatidae*	
Perlidae	*Odontoceridae*	

The ICM index fulfils the requirements of the WFD normative definitions because each criterion is addressed by 2 or 3 of the metrics combined in the ICMi.

- The change in taxonomic composition is mainly evaluated through: number of taxa, EPT taxa,
- The diversity is evaluated through number of taxa
- Sensitive taxa are mainly evaluated with Iberian ASPT (for organic + Nutrient), EPT-taxa and number of selected sensitive taxa (both mainly accounting for hydro-morphological degradation)

Abundance is not considered in the ICMi but is included in some of the national methods (e.g. saprobic index). Additionally an ICMi*quantitativ* including metrics based on quantitative data was calculated.

Most metrics respond to general degradation or combined stressors (Table 7).

Table 7: Relationship between the metrics of the ICMi and stressor indicated by them. Number of crosses indicate the strength of the relationship

Metrics	Organic & nutrient	Hydro-morphology	General habitat degradation
Total # taxa	x	x	xx
# EPT taxa	xx	xx	xx
# selected (sensitive) taxa	x	xx	xx
Iberian ASPT	xxx	x	x

2.5.2 ICMialpine - quantitative data

A comparison between the qualitative and quantitative ICM_{alpine} was carried out for Austria, France, Germany and Slovenia. The Alpine GIG used the approach of the Central Baltic GIG, but quantitative metrics were adapted to fit alpine river types.

The ICM index value is calculated by the weighted sum of all the metrics, according to the conceptual group to which they belong, giving the same weight to each of the three groups (Tolerance, Abundance/Habitat, Richness/Diversity) (Table 8).

Table 8: Metrics of the quantitative ICM_{alpine} (at family level) and their weight for the calculation of the weighted sum.

Type	Metric		Weight
Tolerance	$ASPT_{Iberian}$ - 2		0,333
Abundance/Habitat	Log 10 (sel_sens_taxa)	Log (sum abundance of selected sensitive families)	0,266
	RETI	Rithron feeding type index	0,067
Richness & Diversity	Total number of taxa		0,167
	Number of EPT-taxa		0,083
	Shannon-Wiener Diversity Index		0,083

The conformity of the intercalibration process with the normative definitions of the WFD concerning the abundance is shown by correlating the ICMi*qualitative* to the ICMi*quantitative*. Additionally, the analysis used for the ICMi*qualitative* was also done for the ICMi*quantitative* and is included in the Appendix, Figure 22.

2.6. Minimum quality criteria for data

The member states of the Alpine GIG agreed in accordance to the Central Baltic GIG that some minimum quality criteria should be required from the data sets used in the Intercalibration procedure.

Dataset:

- Minimum number of sites: 20 sites covering widest range of quality classes
- reference state compliant to the REFCOND guidance
- Type description for sites
- squared Pearson coefficient between national index and ICMi: >=0.64 and significant relation (at α=0,05)

Assessment method

- Classification compliant to WFD

If these criteria are not met by some member states, nevertheless the intercalibration is carried out. But the results have to be interpreted in the light of these shortcomings, as indicated by the analysis of plausibility (Chapter 4).

2.7. The 6 steps of the ICMi method

The intercalibration of the national boundary values using the ICMi approach involved six steps that are described in detail in the following chapter:

1. Collection of data sets from every member state
2. Definition and calculation of reference
3. Calculation of EQR – values for the ICMi – metrics
4. Calculation of the ICMi value
5. Calculation of regression between national EQR and ICMi – EQR
6. Transformation of national boundary values into ICMi – EQR values

STEP 1. Collection of data sets from every member state

From every member state the necessary qualitative and (if available) quantitative data was collected for performing the comparison of boundary values by using the ICMi approach. The metrics were collected for the taxonomical *family level*. For every member state this data collection consists of:

General information:	stream name, sampling site, sampling date, sample code
National method results:	EQR value from national assessment method, national Ecological Quality Class, Indication of reference sites, national boundary values
Metrics for ICMi$_{qualitative}$:	total number of taxa, number of EPT taxa, number of selected, (sensitive) taxa, ASPT$_{Iberian}$ - 2

Metrics for ICMi$_{quanitative}$: ASPT$_{Iberian}$ – 2, Log 10 (sel_sens_taxa), RETI, total number of taxa, number of EPT-taxa, Shannon-Wiener Diversity Index

STEP 2. Definition and calculation of reference

The participating member states nominated reference sites in consideration of the harmonized and agreed criteria for the selection of reference sites (see Annex B). Separately for each member state the reference value for each metric was calculated as *median of the values at the reference sites*. Therefore, every country has different reference values for its metrics (Table 9).

Table 9: Reference values for the metrics used for the ICMi.

type R- A1	no. fam	no. EPT fam	no. sens fam	IbASPT-2
Austria	33	15	10	4,67
France	24	11	7	3,71
Italy	19	9	6	4,40
Germany	27	16	9	5,07
Slovenia	28	13	9	4,37

type R- A2	no. fam	no. EPT fam	no. sens fam	IbASPT-2
Austria	39	18	12	4,71
France alp	23	11	8	4,35
France pyr	35	16	13	4,44
Spain	23	11	7	4,22
Italy	17	8	5	3,83

Differences are mainly caused by differing sampling methods. Another influence is the selection of the sampling sites, which is an additional source of variability. This Problem is addressed in Chapter 4.4 and is important for the interpretation of the results.

STEP 3. Calculation of EQR – values for the ICMi – metrics

For the metrics from each sample (in the case of *ICMi$_{qualitative}$* this are 4 values for each sample) the absolute values were converted to EQR values with the formula as described in the WFD:

$$EQR = \frac{\text{metric value}}{\text{reference value}}$$

This harmonization procedure converts the metric values to Ecological Quality Ratios that are theoretically ranging between 0 and 1. As the reference value is calculated as median of the values from the reference sites and as few sites may have "better" values than the reference sites, it is possible that some values exceed 1.

STEP 4. Calculation of the ICMi value

The EQR value of the ICMi is calculated by averaging the EQR values of the metrics for the $ICMi_{qualitative}$:

$$\textbf{EQR ICMi}\boldsymbol{qual} = \frac{EQR_{metric1} + EQR_{metric2} + EQR_{metric3} + EQR_{metric4}}{4}$$

For the calculation of the $ICMi_{quantitative}$ the value of the ICMi is obtained by multiplying each EQR value of the metric with the weight of the metric (given in Table 8) and then calculating the sum:

$$\textbf{EQR ICMi}\boldsymbol{quant} = EQR_{metric1} \text{ x } weight_{metric1} + EQR_{metric2} \text{ x } weight_{metric2} + EQR_{metric3} \text{ x } weight_{metric3} + EQR_{metric4} \text{ x } weight_{metric4} + EQR_{metric5} \text{ x } weight_{metric5} + EQR_{metric6} \text{ x } weight_{metric6}$$

This calculation is done for the data from each sample from each member state.

STEP 5. Calculation of regression between national EQR and ICMi – EQR

For each member state the EQRs from the national assessment method are correlated with the corresponding EQRs from the ICMi. The regression is graphically illustrated and the regression formula and R² are calculated. Figure 2 shows the regression for Austria for the type RA-1.

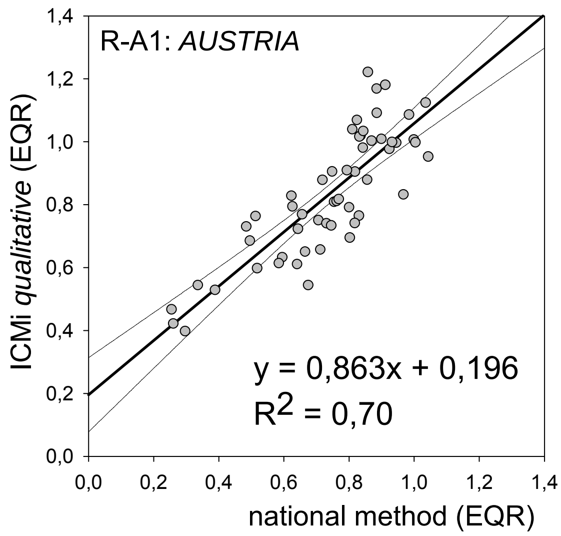


Figure 2: Regression between EQR of the national method and EQR from the ICMi *qualitative* for Austria, IC-type RA1.

STEP 6. Transformation of national boundary values into ICMi – EQR values
With the regression the national boundaries can easily be transformed into ICMi values. Figure 3 illustrates the transformation process of the GOOD/MODERATE boundary for the Austrian dataset of type R-A1. Additionally the 95% confidence limits of the ICMi value that corresponds to the national boundary can be read from the graph.

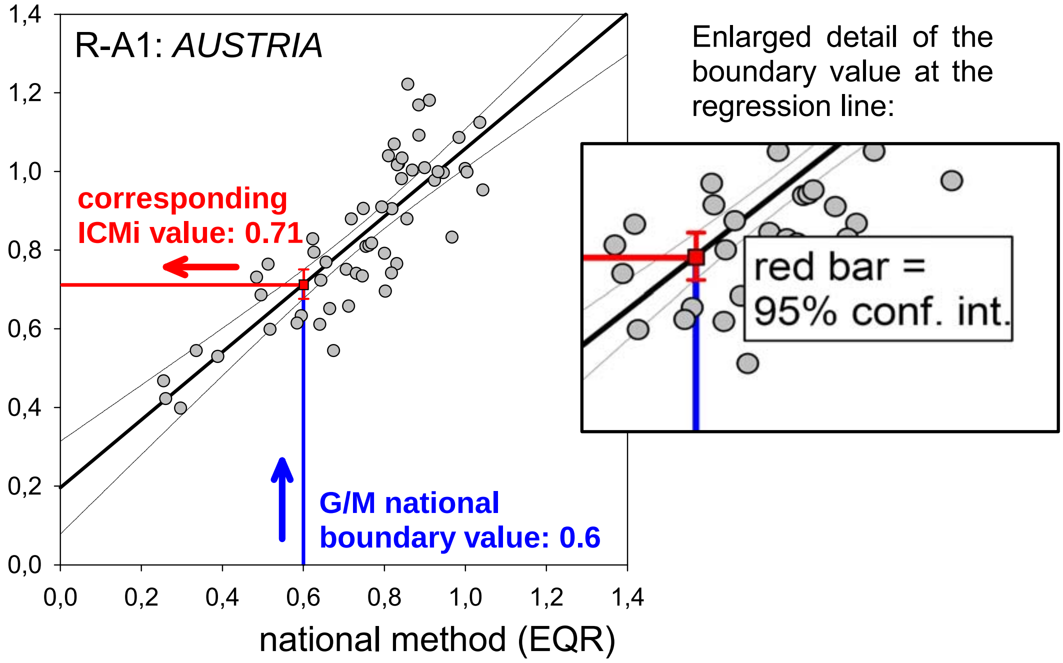


Figure 3: Transformation of the Austrian GOOD/MODERATE national boundary value into an ICMi value +/- 95% confident limits, using the regression shown in Figure 2.

In the Alpine GIG Intercalibration process the real transformation and the calculation of the 95% confidence limits was done by calculations using the regression formulae of STEP 5.

2. 8. The accepted variation of the boundary value

Due to several sources of variability (for a more detailed analysis concerning the Alpine GIG dataset see chapter 4) a recalculation from an ICMi boundary value to a national EQR value would not give exactly the national boundary value – especially when a dataset other then the one of the intercalibration exercise is used.

As a consequence the Alpine GIG suggests to use an „*acceptable range of variation*“ rather than a fixed value alone.

As value for this „*acceptable range of variation*" the GIG proposes ¼ of the median status class width of the participating member states.
Background of this proposal: The results of the assessment methods are subject to several sources of variation. Thus the status assessment is somehow more significant in the middle of a status class than compared to the transitional zone to the neighbouring status classes. This "insecure" zone of assessment is assumed to be ¼ of the status class width (more detailed estimates of accuracy and precision is lacking in most countries at the moment). Figure 4 visualizes this proposal for a „*acceptable range of variation*".

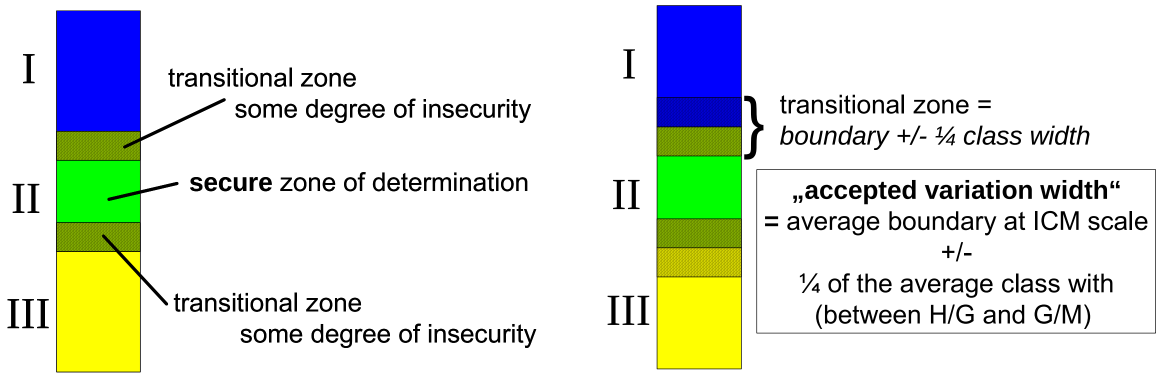


Figure 4: **Left**: The first three status classes: High, Good and Moderate. Every status class can be segmented in an area where the status evaluation is secure (high confidence) and in transitional zones to the neighbouring status classes, where the status evaluation is to some extent insecure. **Right**: The transitional zone between two status classes is supposed to the "accepted with of variation".

3. Results of the Alpine Geographical Intercalibration Group

3.1. Correlation between national method and ICMi *qualitative*

The correlation between the EQR values of the national methods and both the single ICMi-metrics and the $ICMi_{qualitative}$ and $ICMi_{quantitative}$ is shown in the Figures 11 to 20 in the Appendix for both intercalibration types and all member states, together wit the regression formulae and the values for R^2.

Figure 5 shows the results of the 6 Step Intercalibration procedure for both intercalibration types. The corresponding values are given in Table 10.

Table 10: Result of the intercalibration procedure with the $ICMi_{qualitative.}$

MS	National boundary		National boundary ICMi				Proposed range ICMi (+/-)	
	H/G	G/M	H/G value	+/-95% CL	G/M value	+/-95% CL	H/G	G/M
Type R-A1								
Austria	0,80	0,60	**0,89**	***0,03***	**0,71**	***0,04***	0,82 - 0,92	0,66 - 0,76
France	0,93	0,79	**0,89**	***0,03***	**0,74**	***0,04***	0,82 - 0,92	0,66 - 0,76
Germany	0,80	0,60	**0,87**	***0,05***	**0,76**	***0,03***	0,82 - 0,92	0,66 - 0,76
Italy	0,91	0,72	**0,81**	***0,03***	**0,56**	***0,03***	0,82 - 0,92	0,66 - 0,76
Slovenia	0,80	0,60	**0,84**	***0,13***	**0,66**	***0,13***	0,82 - 0,92	0,66 - 0,76
Type R-A1								
Austria	0,80	0,60	**0,82**	***0,02***	**0,65**	***0,02***	0,77 - 0,87	0,58 - 0,68
France alp	0,93	0,71	**0,89**	***0,03***	**0,63**	***0,04***	0,77 - 0,87	0,58 - 0,68
France pyr	0,94	0,81	**0,80**	***0,02***	**0,64**	***0,03***	0,77 - 0,87	0,58 - 0,68
Spain	0,83	0,51	**0,81**	***0,01***	**0,56**	***0,03***	0,77 - 0,87	0,58 - 0,68
Italy	0,94	0,75	**0,85**	***0,03***	**0,59**	***0,04***	0,77 - 0,87	0,58 - 0,68

The bands for the "*proposed range of the ICMi*" represent the median boundary values of all member states +/- ¼ of the median good status class range of all member states, as described in chapter 2.8. Table 11 shows the calculation of this value.

Table 11: Calculation of the width of the bands shown in Figure 10. For description see chapter 2.8.

R-A1	class width G/M
Austria	0,2
France	0,14
Italy	0,19
Germany	0,2
Slovenia	0,2
median class width:	**0,2**
1/4 of class width:	**0,05**

R-A2	class width G/M
Austria	0,2
France alp	0,22
France pyr	0,13
Spain	0,32
Italy	0,19
median class width:	**0,2**
1/4 of class width:	**0,05**

As an alternative approach the bands of "*accepted variation*" were also calculated as median confidence limits of all member states – to be found in the Appendix, Figure 21. From the statistical point of view this approach does not seem to be appropriate, as the confidence limits are decreasing with increasing data quality, but the sources of variation if the results are still present – see Chapter 4).

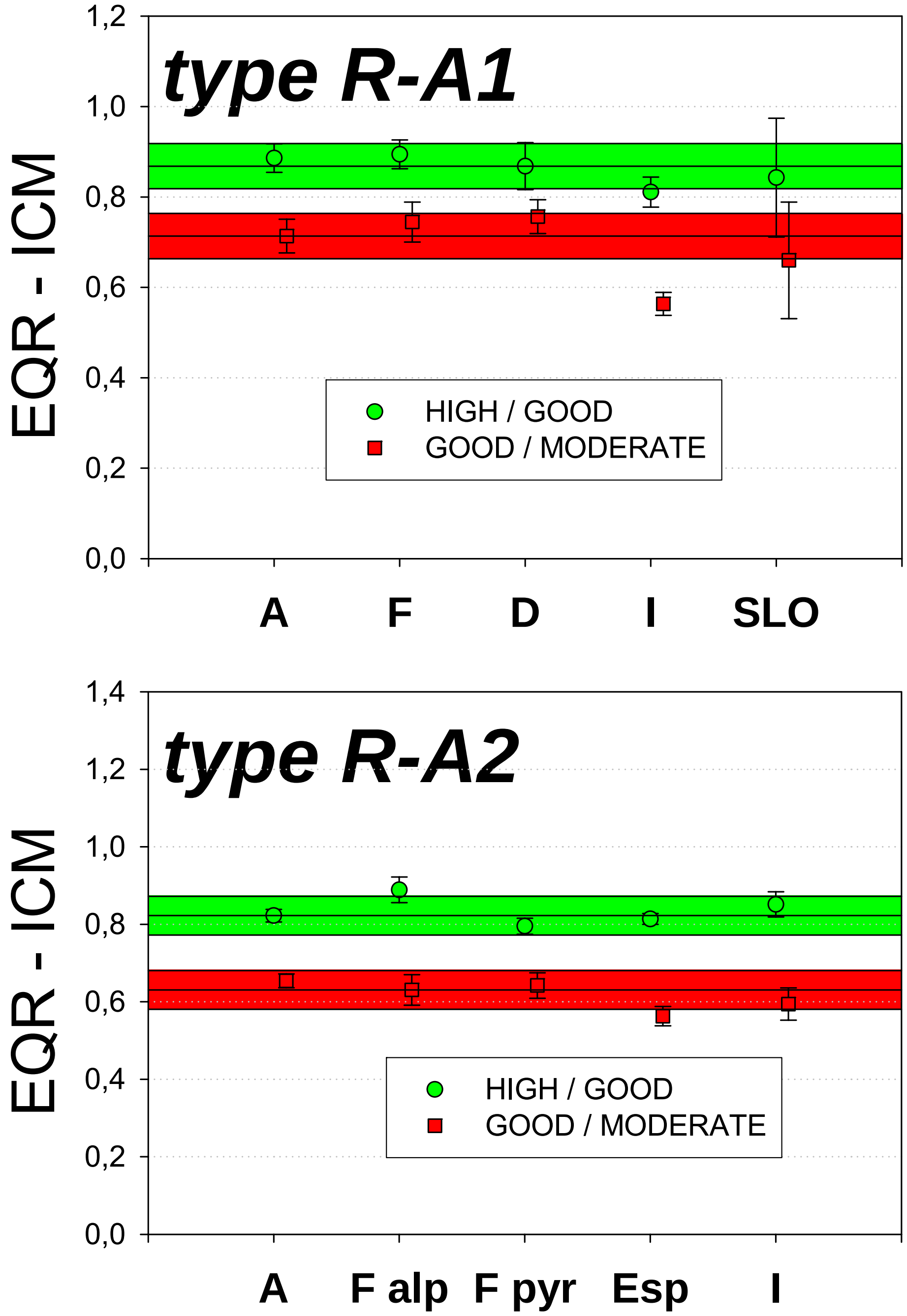
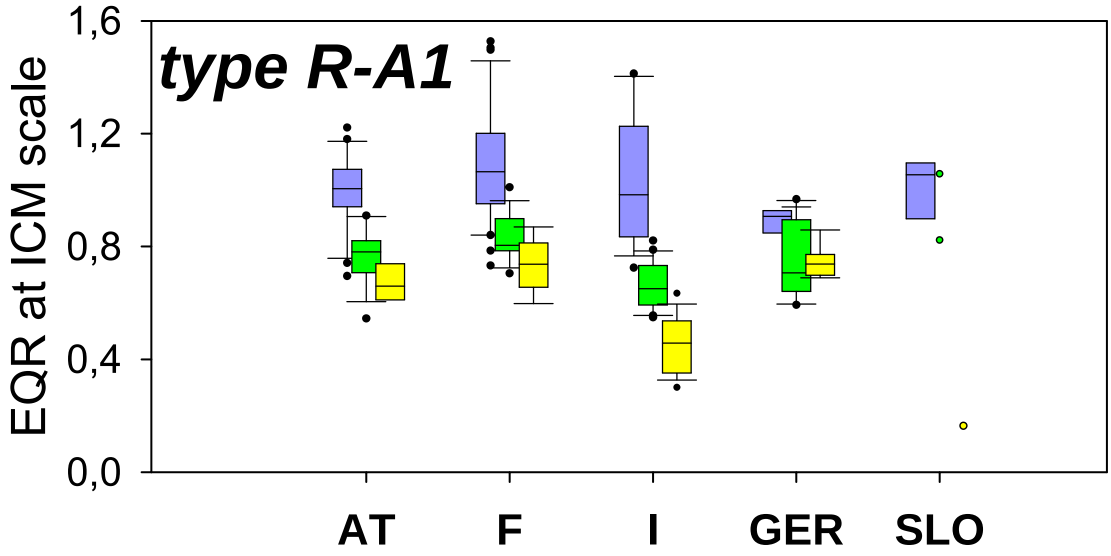


Figure 5: Boundary values of the member states for the Alpine GIG types as calculated from the regression between national method and ICMi values. Given are the values for the high/good and the good/moderate boundary and the 95% confident limits, expressed in EQR of the ICMi method. The bands represent the median boundary values of all member states +/- ¼ of the median good status class range of all member states.

3.2. Discriminatory power of the ICMi class boundaries

Figure 6 indicates the status classes HIGH, GOOD and MODERATE for the two types and for all member states. The discriminatory power of the ICMi is limited due to the various sources of variation that impact the correlation between the national method and the ICMi, as described in Chapter 4. However, the separation of the three status classes and the accordance between the member states is very good for type R-A2. For type R-A1 there are some problems, especially for the German and the Slovenian data set – based on data quality limitations (see 4.1.).



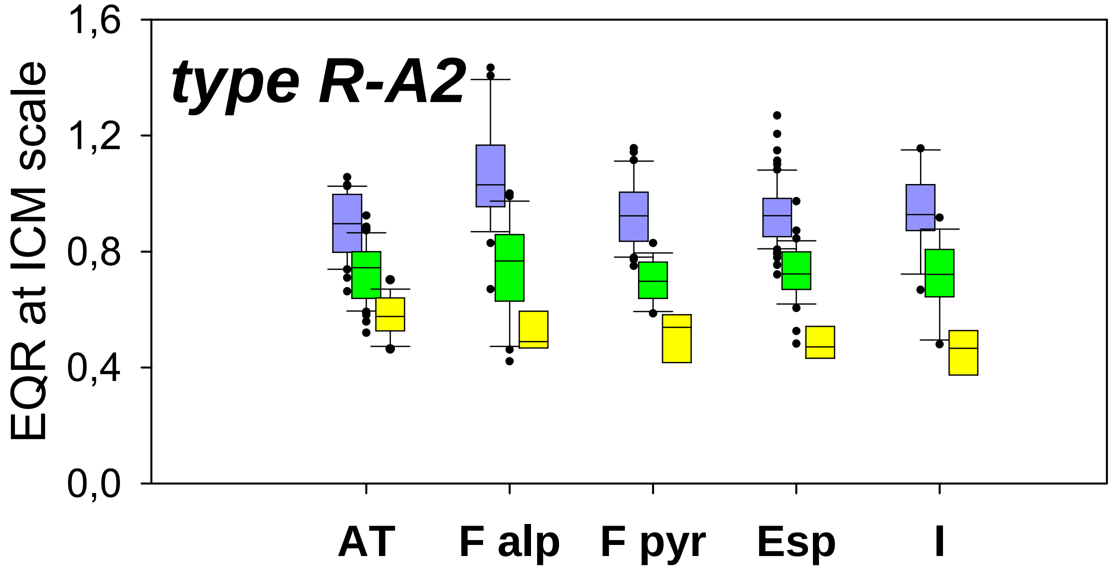


Figure 6: Ranges of status classes (national classification) expressed with ICMi EQR values. Boxplots: boundary of the box closest to zero indicates the 25th percentile, a line within the box marks the median, and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above and below the box indicate the 90th and 10th percentiles. Points indicate outliers. For Slovenia status class II and III is represented by 2 and 1 data point, respectively.

3.3. Correlation between ICMi$_{qualitative}$ and ICMi$_{quantitative}$

There is a strong correlation between the qualitative and the quantitative ICMi methods (Figure 7), indicating that both approaches result in a consistent intercalibration of the class boundaries.

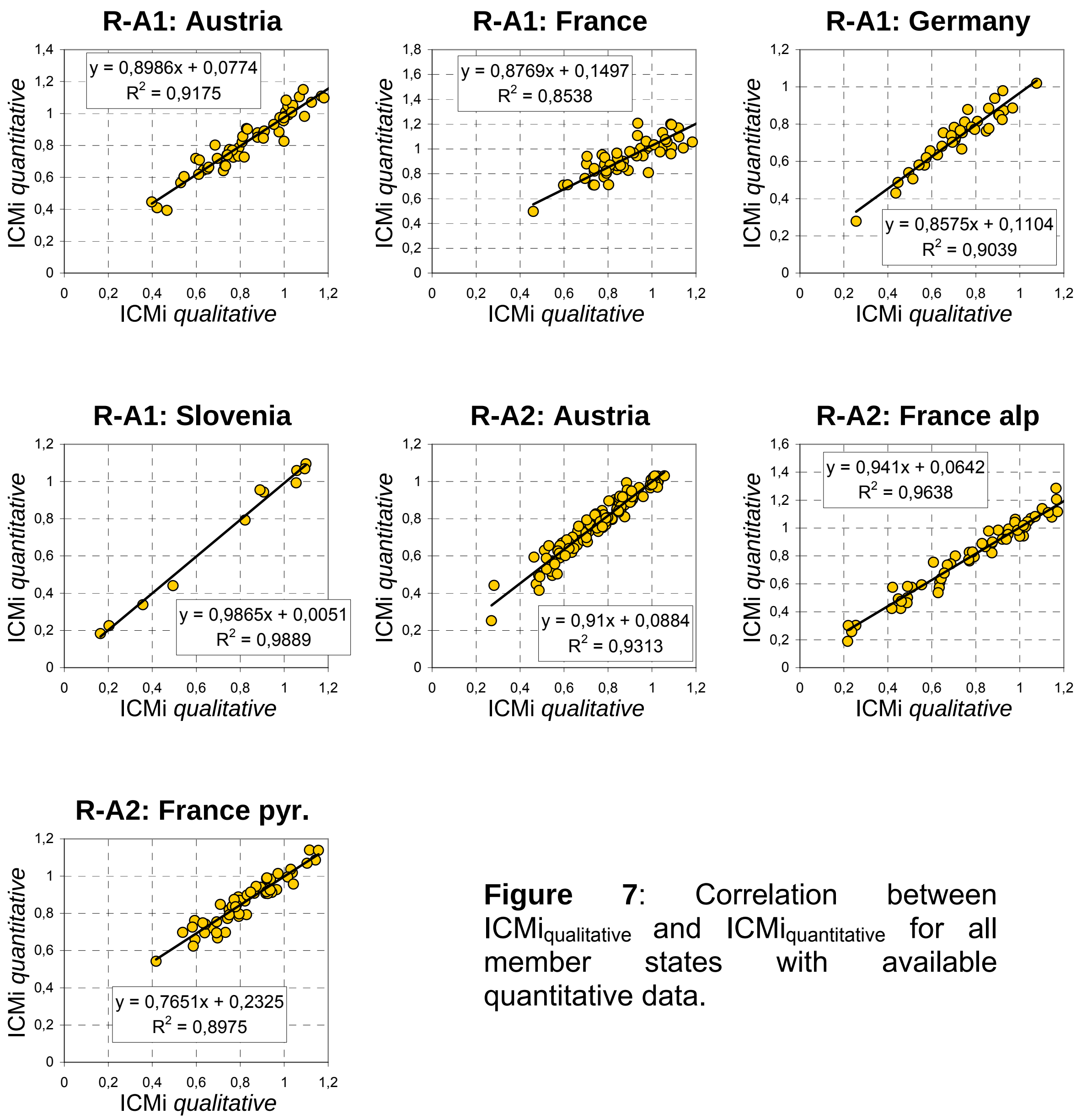


Figure 7: Correlation between ICMi$_{qualitative}$ and ICMi$_{quantitative}$ for all member states with available quantitative data.

The intercalibration analysis accomplished with the ICMi$_{quantitative}$ is included in Appendix III.

4. Discussion of the variability and plausibility of the results

As the boundary values at the ICMi scale are obtained by transforming the national boundary values using the regression formula there are several factors that add variation to the results, in particular data limitations, natural variability and the simplification principles of the ICMi.

4.1. Data limitations

The data submitted by the member states is limited in quantity and quality. Therefore the regression formula will change slightly with new data sets. In this context the lack of bad quality sites is a problem, as in some cases few bad quality sites have a distinct influence on the slope of the regression (Figure 8). For some member states bad quality sites are missing completely. The minimum data quality criteria are compiled in Table 12

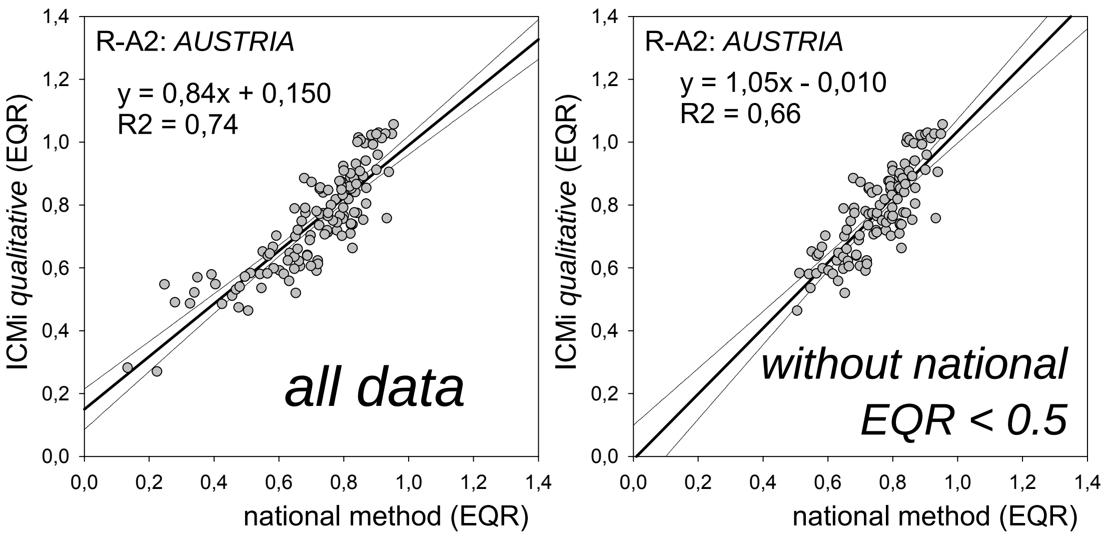


Figure 8: Regressions as described in STEP 5. Left: Austria type R-A2, complete data set. Right: Same data, but points with national EQR < 0.5 are missing.

Table 12: Minimum data quality criteria.

	A	GE	E	F	I	SLO
20 sites covering widest range	√	√	√	√	√	
status class IV and V is missing	√	√	√	**for A1 and A2 pyr**	**for A2 only**	√
reference state compliant to REFCOND	√	√	√	√		√
type description for sites	√	√	√	√		√
R^2 >=0.64 (at a=0,05)	√	√	√	√	√	√
no differences between EQR of reference sites for national method and ICMi (see Figure 10)	**for A2 only**		√	**for A2 pyr only**	√	√
national method compliant to WFD	√	√	√	√		√

4.2. Natural variability

Biological systems are generally characterized with a high degree of variability, and in particular this is the case for benthic invertebrates in running waters, as has been shown by numerous scientific studies. Thus, two investigations of ecological status would hardly result in exactly the same index value – the Water Framework Directive considers this with the concept of accuracy and precision.

The evident consequence of this variability is that the discriminatory power between the status classes is reduced when ICMi values are used (This is also the reason why the ICMi metrics can not substitute the national methods). This means just that the resolution of the results is not as fine at the ICM – scale when compared to the national scale. Figure 9 shows this comparison of status classes with national EQRs and ICMi EQRs.

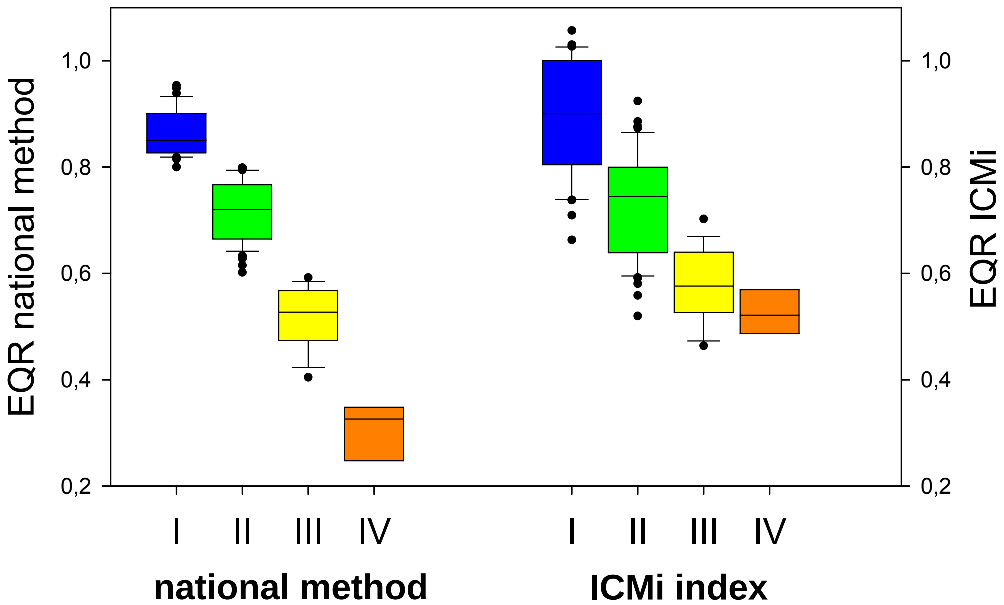


Figure 9: Ranges of status classes (data from Austria Type R-A2) expressed with national EQR values (left) and ICMi EQR values (right). Boxplots: he boundary of the box closest to zero indicates the 25th percentile, a line within the box marks the median, and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above and below the box indicate the 90th and 10th percentiles. Points indicate outliers.

4.3. Simplification principle of the ICMi

When interpreting the results of the intercalibration process we should have in mind that the ICMi method is a simplifying approach for a comparison at a European scale. In contrast to the ICMi the national methods have a finer resolution in terms of typology and data quality. For the ICMi we use a simplified typology and a high taxonomic level (family level). This is just an additional argument for accepting a certain range of results instead of a single boundary value.

4.4. The inter-relationship: reference values – ICMi results

The reason for the harmonization procedure of the national data and boundary values (as described in chapter 2.7) is to enable an intercalibration of the various boundary values at a comparable level.
Due to the EQR – harmonization procedure (STEP 3) the reference values of the ICMi metrics play a crucial role in the interpretation of the data:
Differences between member states can either be caused by differences in setting the class boundary values, or by differences in the interpretation of reference conditions and the selection of reference sites.

The median EQR of the ICMi method of the reference sites is 1 by definition (see STEP 2). In general the median EQR of the national assessment method should also be very close to 1 because the reference criteria stated in Annex B should be valid also on the national level. Obviously this is not the case for some member states (Figure 10).

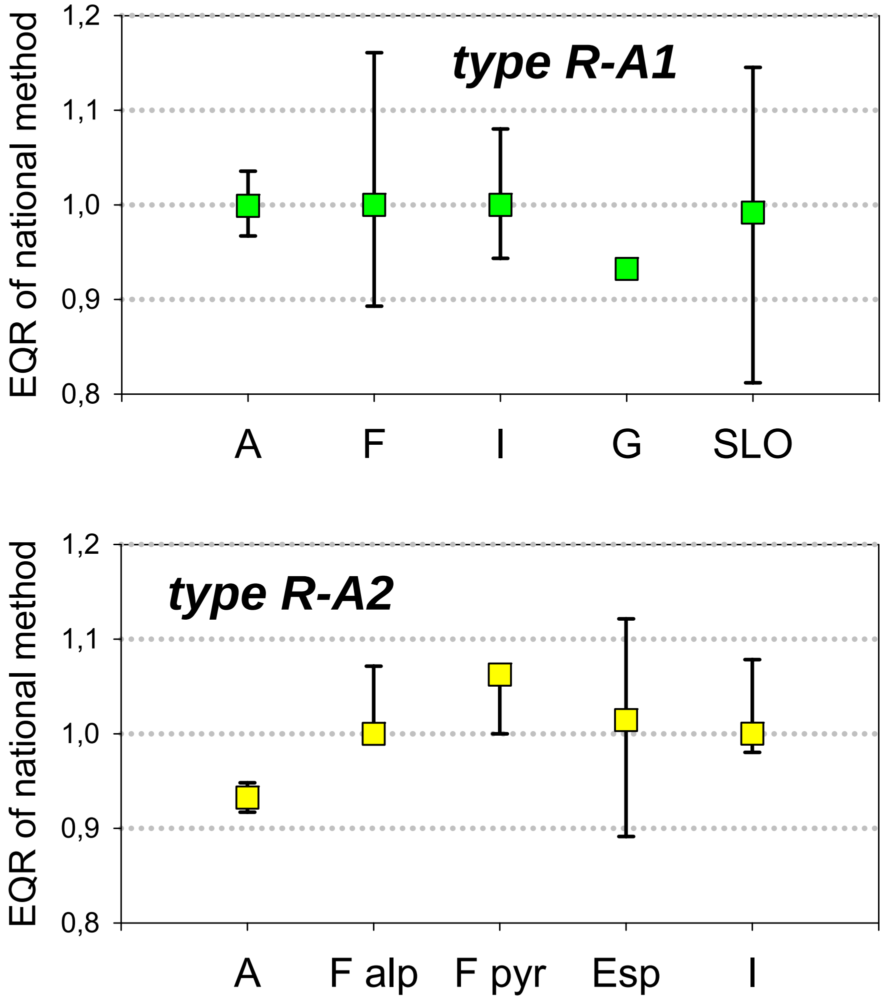


Figure 10: EQR values from the national method for the reference sites (for number of sites see Table 4): 25th percentile – median – 75th percentile.

These differences between the reference values at the ICMi and at the national scale can have several reasons:

a) Reference sites used for the ICMi are not identical with the reference sites used for the national approach. One potential reason for that could be that the submitted data file for the intercalibration is somehow limited (e.g. there are simply no sites included with an EQR of 1).
b) The reference sites are the same as for the national method but the reference values for the national method were obtained with a different approach than the (simplified) ICMi – method. (This is the case e.g. for Austria where the calculation of the reference value involved the 75^{th} percentile of the reference sites and the values are re-normalized in a way that identical boundary values can be used for all national types.)

This influence of the reference sites on the result of the intercalibration is difficult to overcome. The GIG intensely discussed this problem in detail and came to the conclusion that a further harmonization (e.g. setting the EQR for the reference at both scales to the same value) would not solve this problem: the graphs would only indicate the deviation of the ICMi values from a harmonized reference value, without taking into account differences in setting the national reference.

However the influence of the reference site selection should be taken into account for the interpretation of the results: when the EQR of the reference sites at the national level is below 1 this means that a national EQR < 1 corresponds to a ICMi – EQR of 1. Thus in this case all ICMi values are somewhat increased (see also Table 12).

Full comparability is only given, when the median national EQR of the reference sites is close to 1.

When comparing boundary values of two member states at the ICMi scale there can be two reasons for remarkable differences:

- *differences in the national boundary setting (focus of the intercalibration!)*
- *differences in selection of the reference sites*

For the interpretation of the results we can use a simple relationship:

- ***national EQR of reference sites < 1 → → → higher ICMi values***
- ***national EQR of reference sites > 1 → → → lower ICMi values***

4.5. Overall statement to the variability of the results

Taking into account the arguments discussed in the chapters 4.1. – 4.4. the result of the intercalibration for both alpine stream types seem to be in an acceptable range, especially for the type R-A2. In the type R-A1 Italy is not within the proposed range for the boundary values – this could be subject to a discussion of the national boundary values. The Slovenian data is characterized by high variability – caused by the low number of sites and probably typological reasons (streams have a karstic characteristic and very small catchments). But this problem can be solved in future when more comprehensive data collections will be available.

5. GIG members – contact information

MS	Name	Institution	Email
Austria	Gisela Ofenböck	Ministry of Water Management/Austria	gisela.ofenboeck@lebensministerium.at
Austria	Franz Wagner	BAW/Austria	franz.wagner@baw.at
France	Jean-Gabriel Wasson	Cemagref Lyon, France	Jean-gabriel.wasson@cemagref.fr
France	Nicolas Mengin	Cemagref Lyon, France	mengin@lyon.cemagref.fr
Germany	Ilona Schlößer	Bayer. Landesamt für Umwelt	Ilona.schloesser@lfu.bayern.de
Germany	Folker Fischer	Bayerisches Landesamt für Wasserwirtschaft	Folker.Fischer@lfw.bayern.de
Italy	Paolo Negri	APPA, Italy	Paolo.negri@provincia.tn.it
Italy	Maurizio Siligardi	APPA, Italy	Maurizio.siligardi@provincia.tn.at
Slovenia	Gorazd Urbanic	Institute for water of the Republic of Slovenia	Gorazd.urbanic@bf.uni-lj.sl
Slovenia	Bernarda Rotar	Environmental Agency of the Republic of Slovenia	Bernarda.rotar@gov.si
Spain	Miriam Pardos	Confederacion Hidrografica del Ebro, Spain	mpardos@chebro.es
Spain	Concha Duran	Confederacion Hidrografica del Ebro, Spain	cduran@chebro.es

6. Appendix

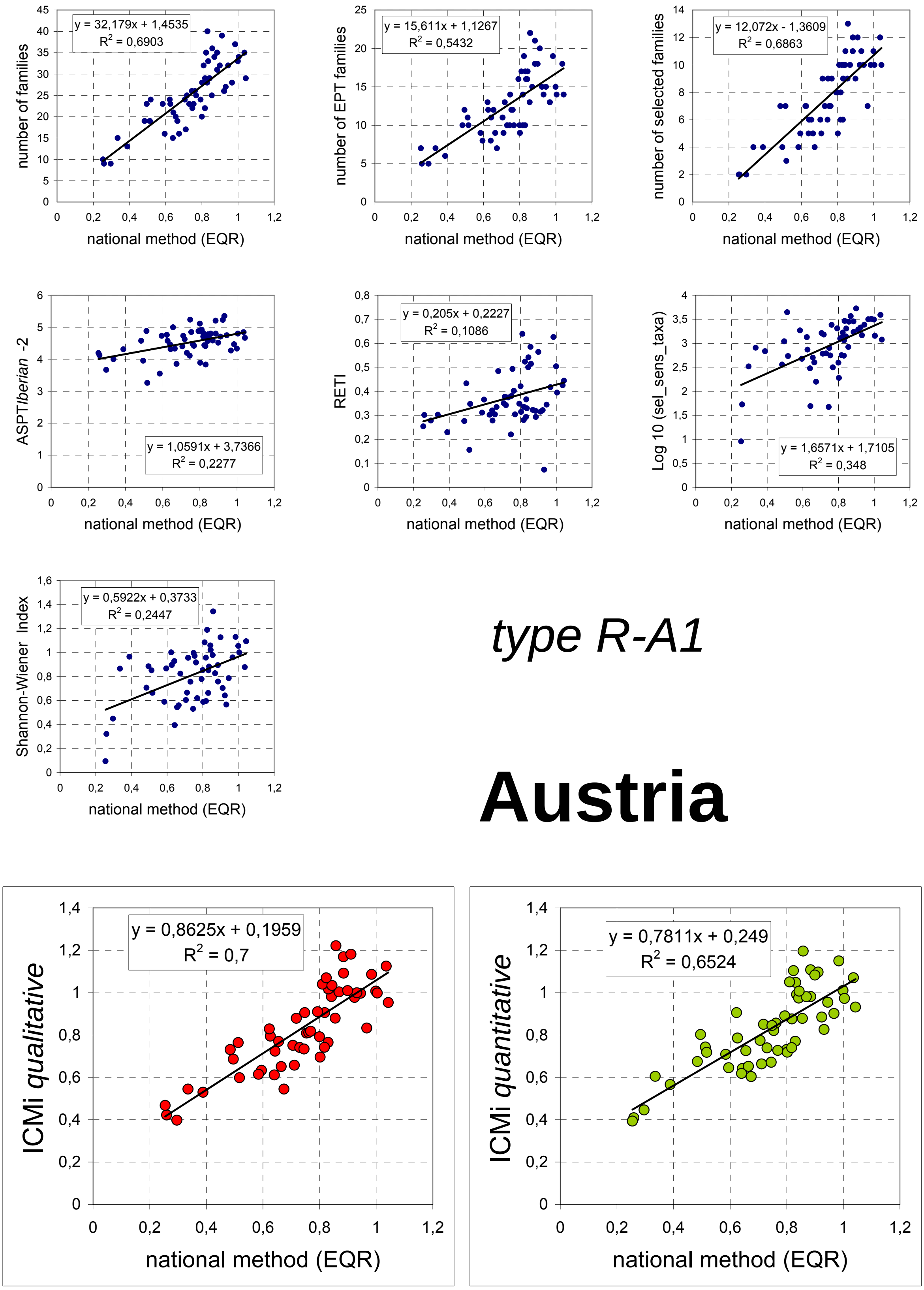


Figure 11: Correlation between the EQR values from the national methods and the EQR values from the single metrics as well as the ICMi *qualitative* and the ICMi *quantitative*.

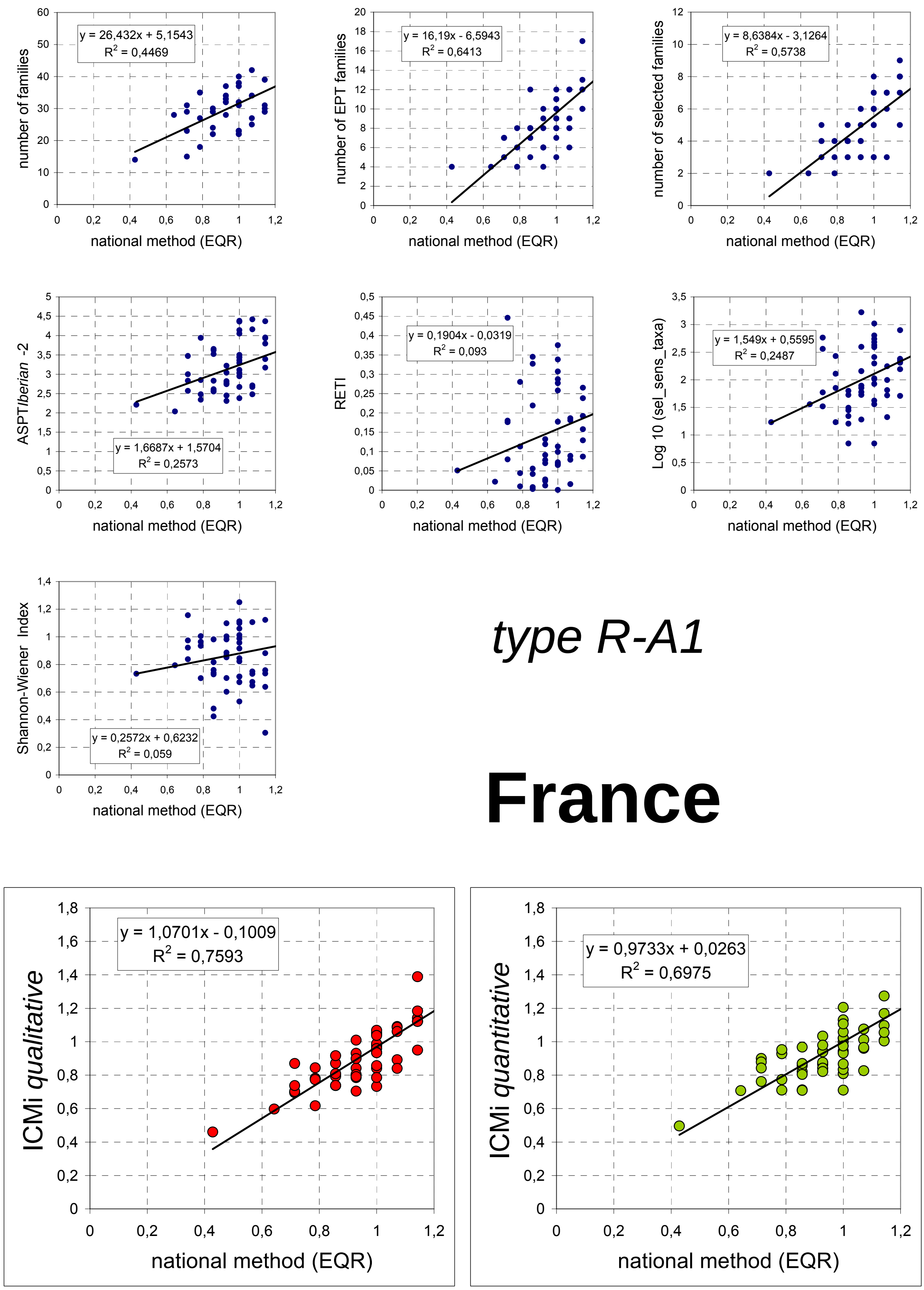


Figure 12: Correlation between the EQR values from the national methods and the EQR values from the single metrics as well as the ICMi *qualitative* and the ICMi *quantitative*.

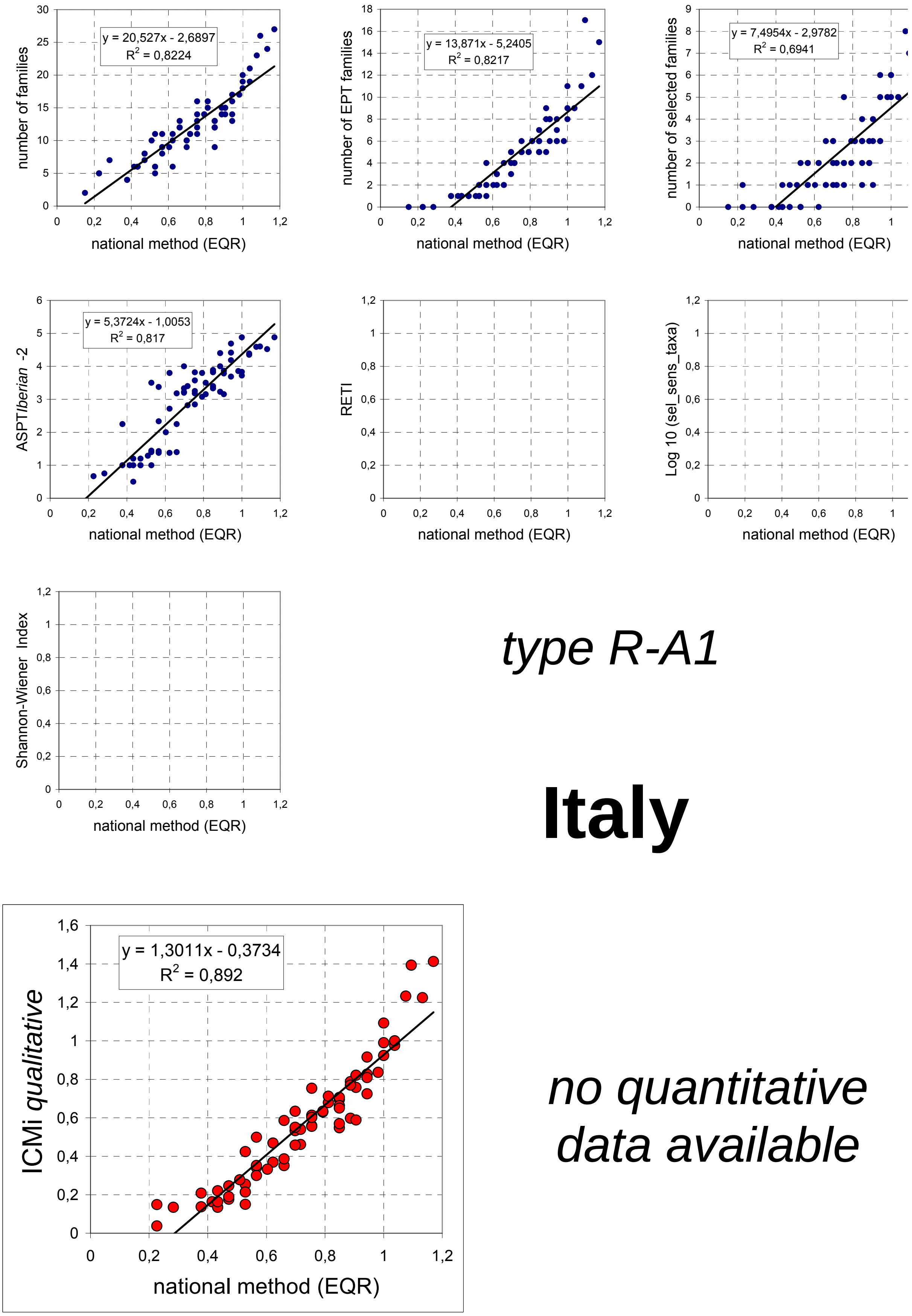


Figure 13: Correlation between the EQR values from the national methods and the EQR values from the single metrics as well as the ICMi *qualitative* and the ICMi *quantitative*.

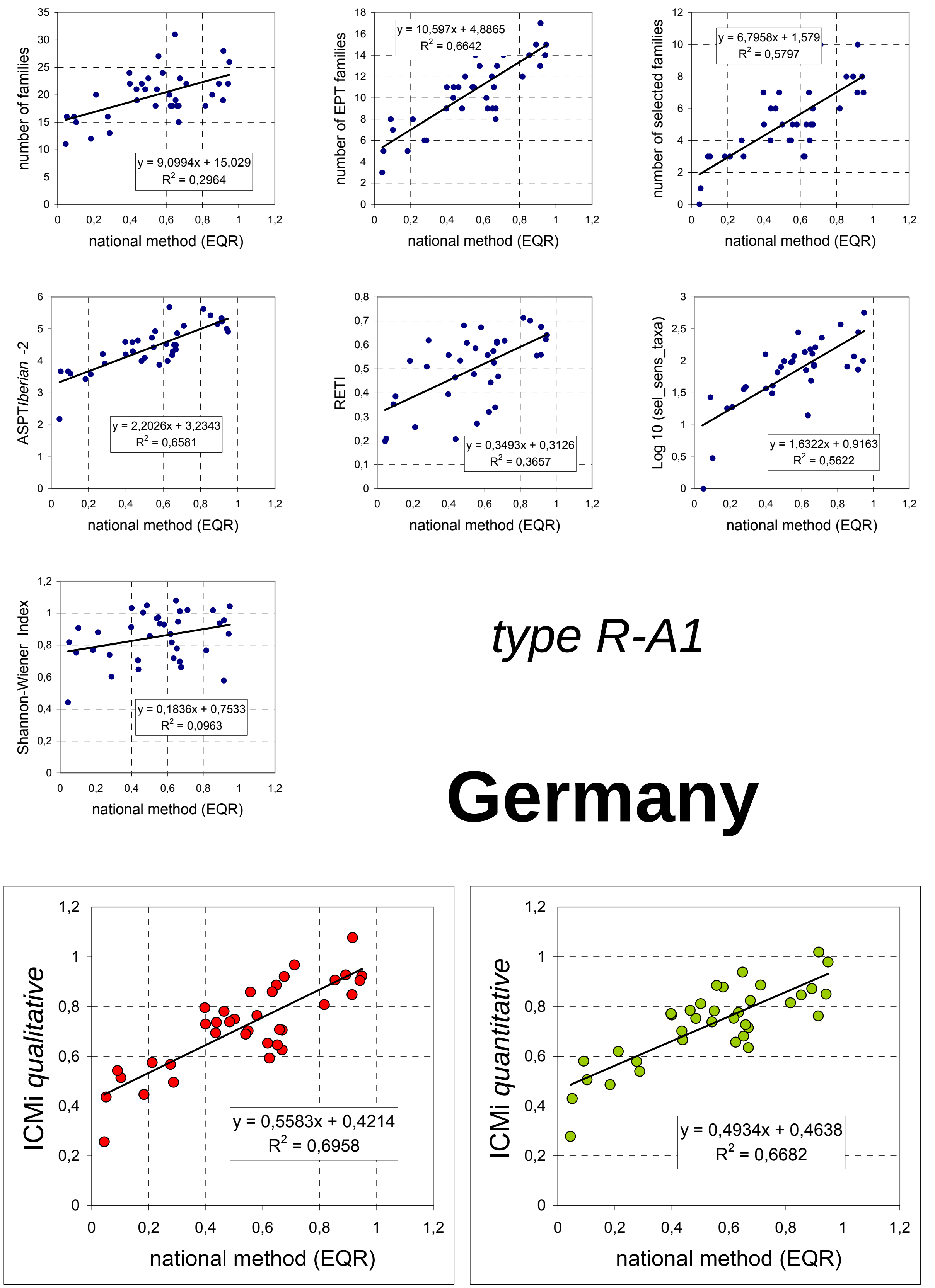


Figure 14: Correlation between the EQR values from the national methods and the EQR values from the single metrics as well as the ICMi *qualitative* and the ICMi *quantitative*.

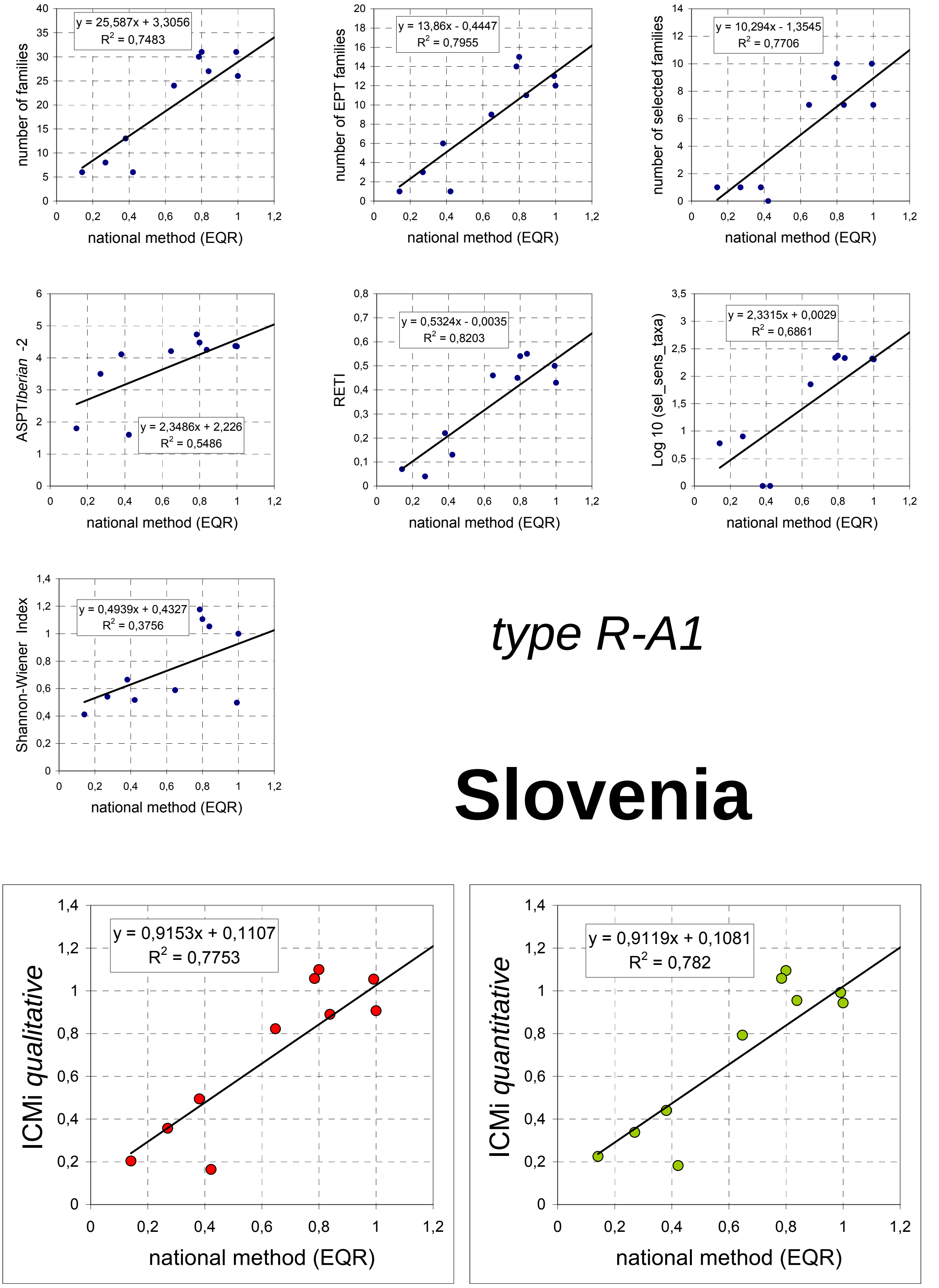


Figure 15: Correlation between the EQR values from the national methods and the EQR values from the single metrics as well as the ICMi *qualitative* and the ICMi *quantitative*.

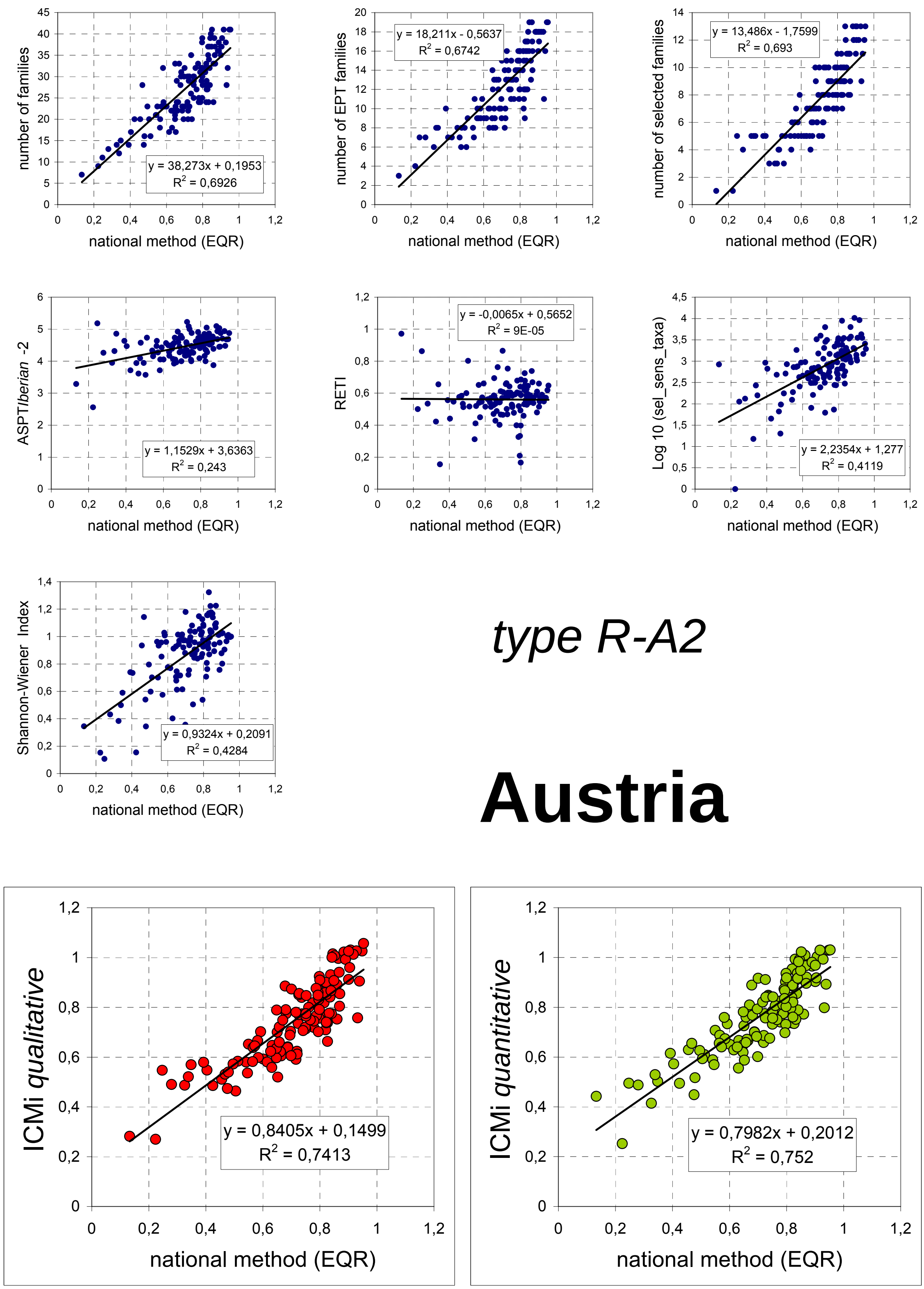


Figure 16: Correlation between the EQR values from the national methods and the EQR values from the single metrics as well as the ICMi *qualitative* and the ICMi *quantitative*.

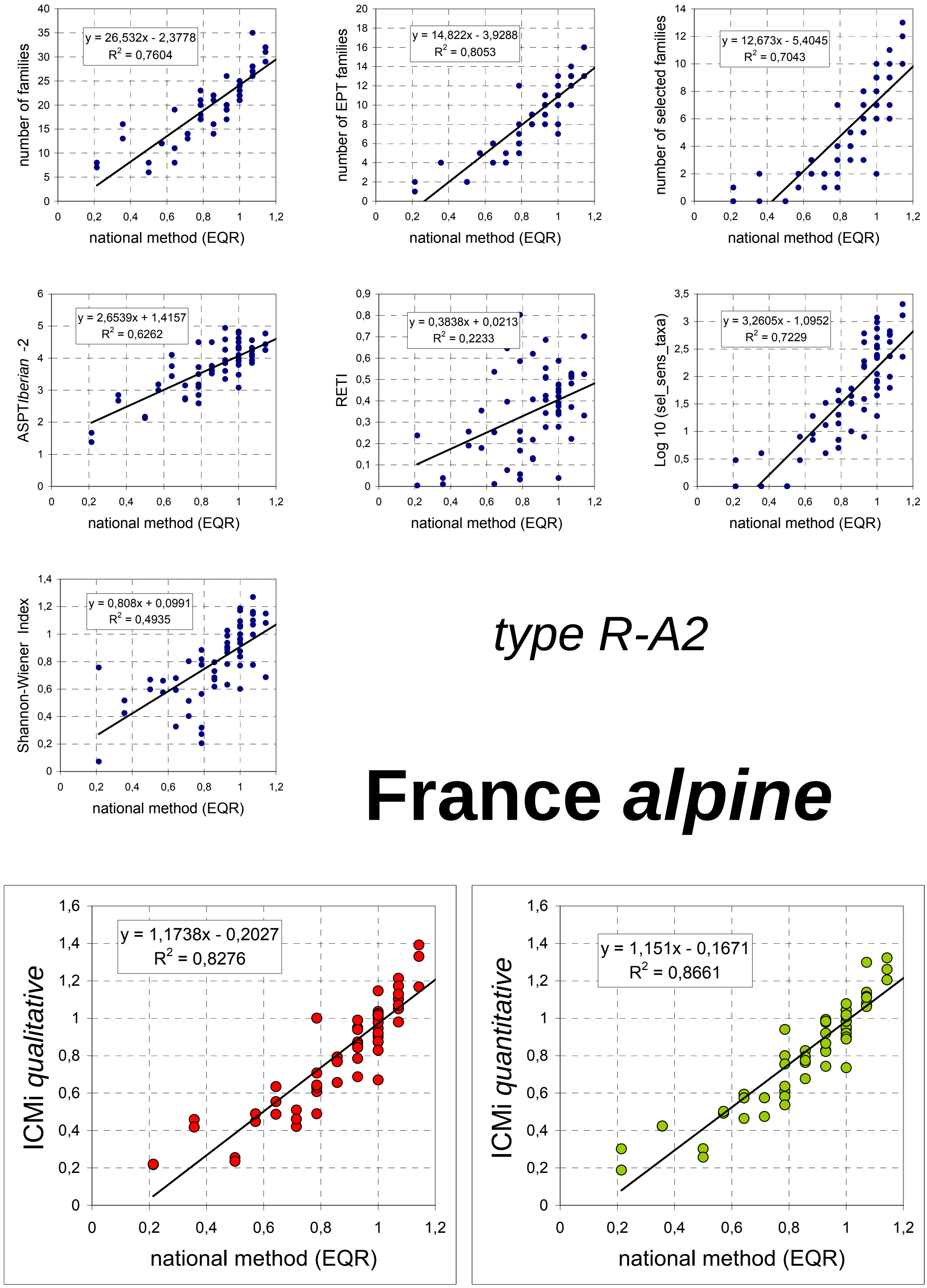


Figure 17: Correlation between the EQR values from the national methods and the EQR values from the single metrics as well as the ICMi *qualitative* and the ICMi *quantitative*.

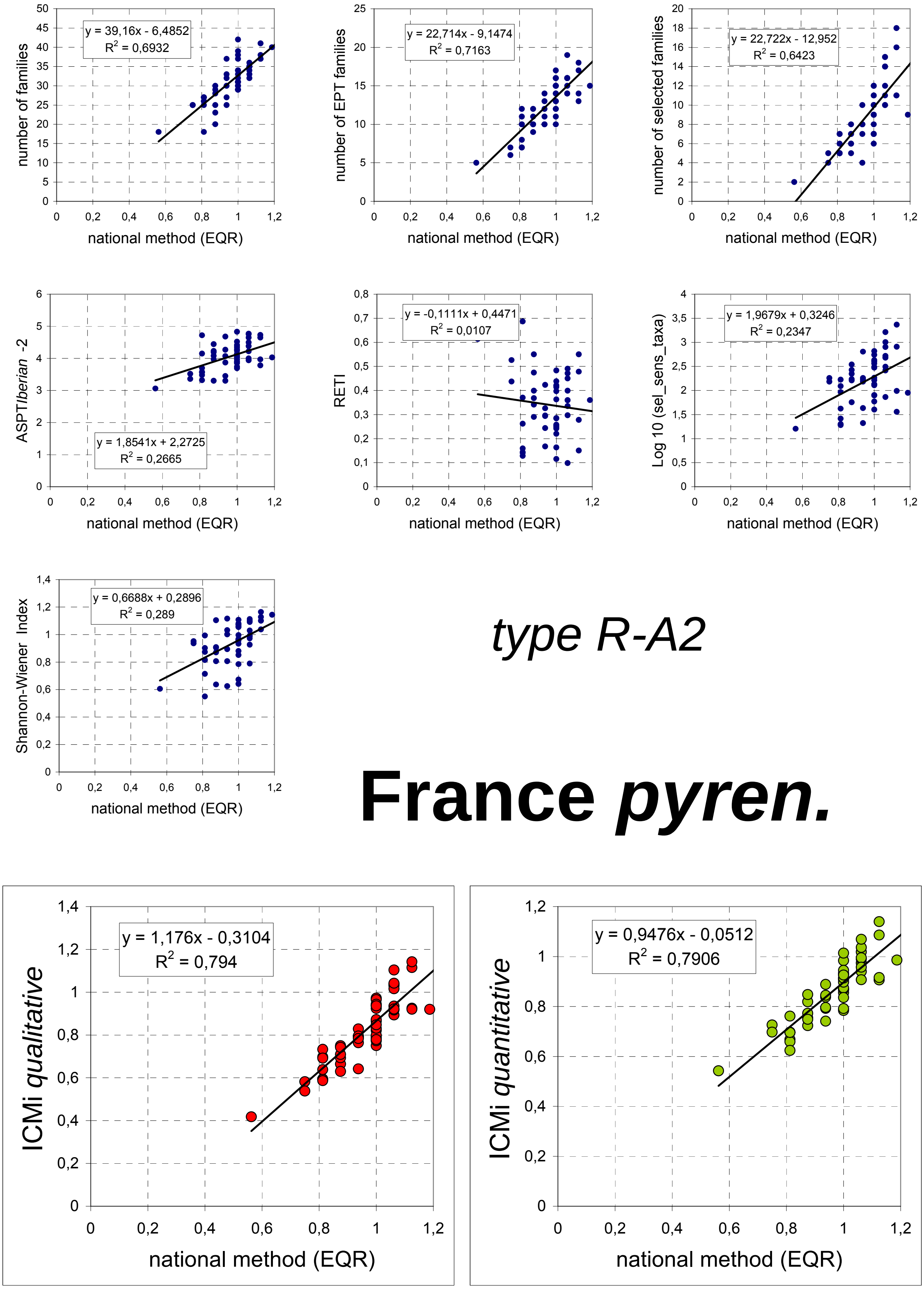
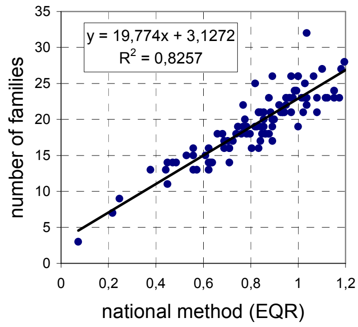
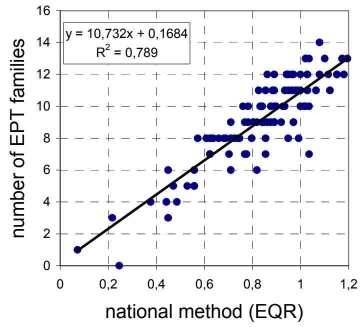
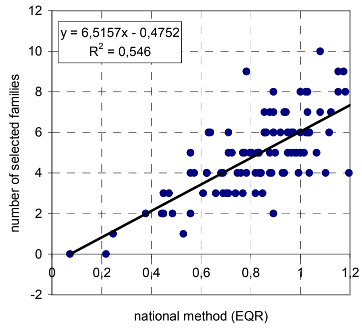
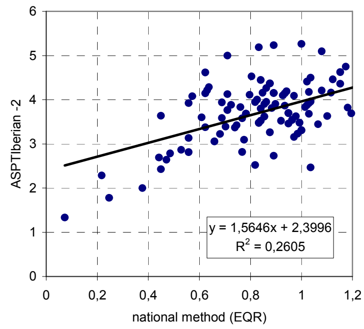


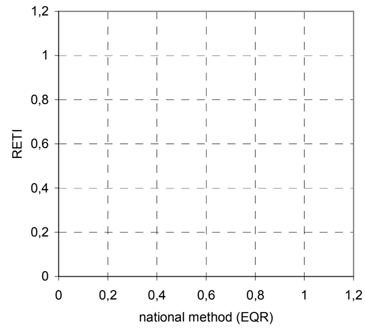
Figure 18: Correlation between the EQR values from the national methods and the EQR values from the single metrics as well as the ICMi *qualitative* and the ICMi *quantitative*.

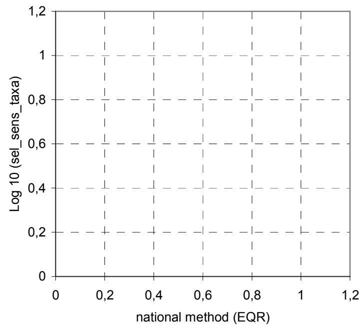


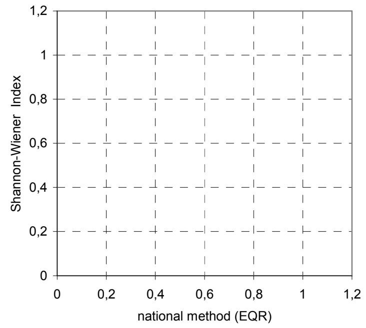






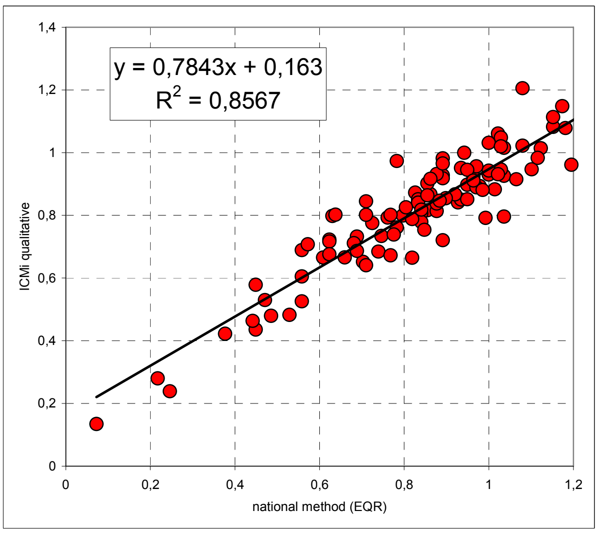






type R-A2

Spain



no quantitative data available

Figure 19: Correlation between the EQR values from the national methods and the EQR values from the single metrics as well as the ICMi *qualitative* and the ICMi *quantitative*.

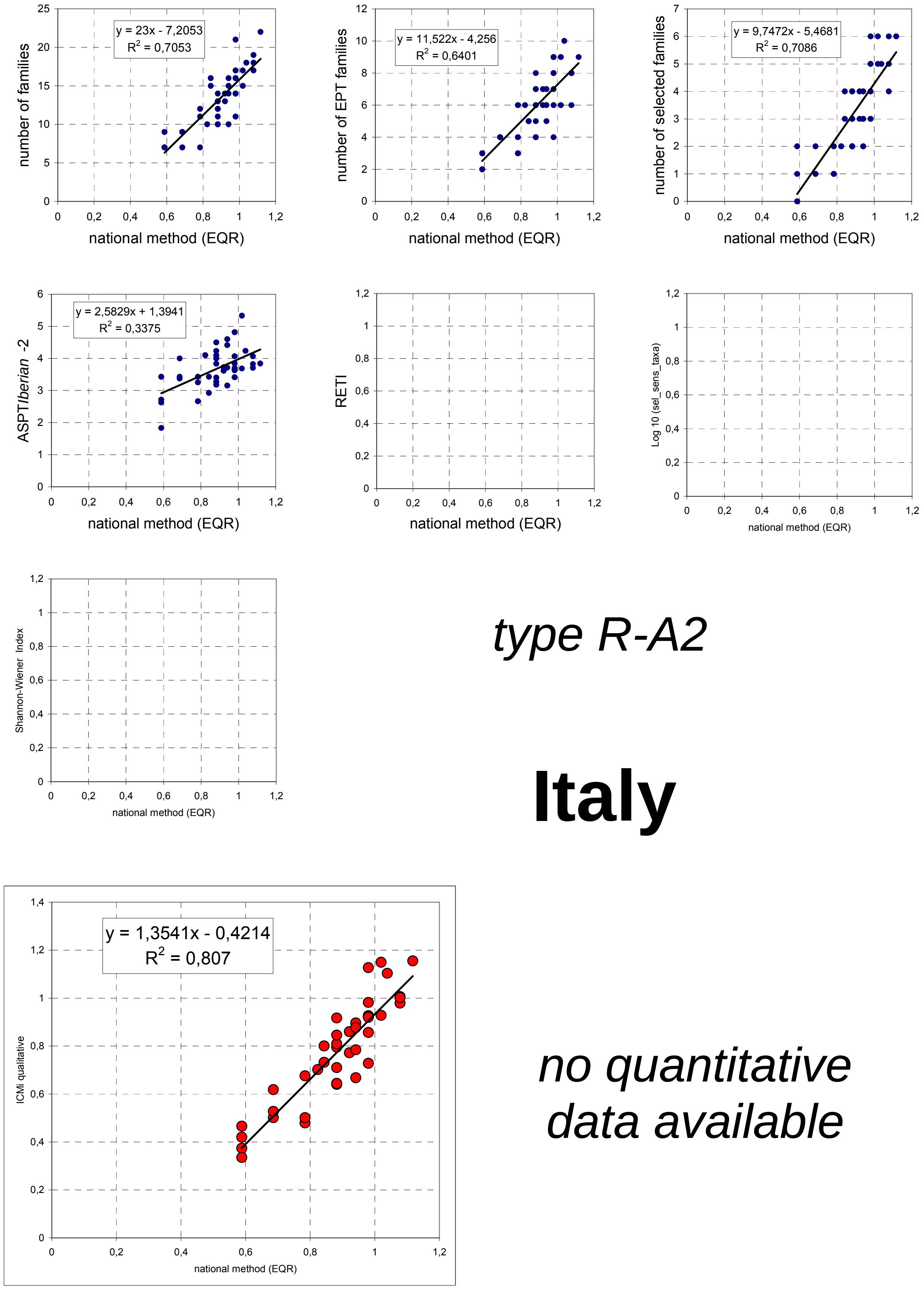
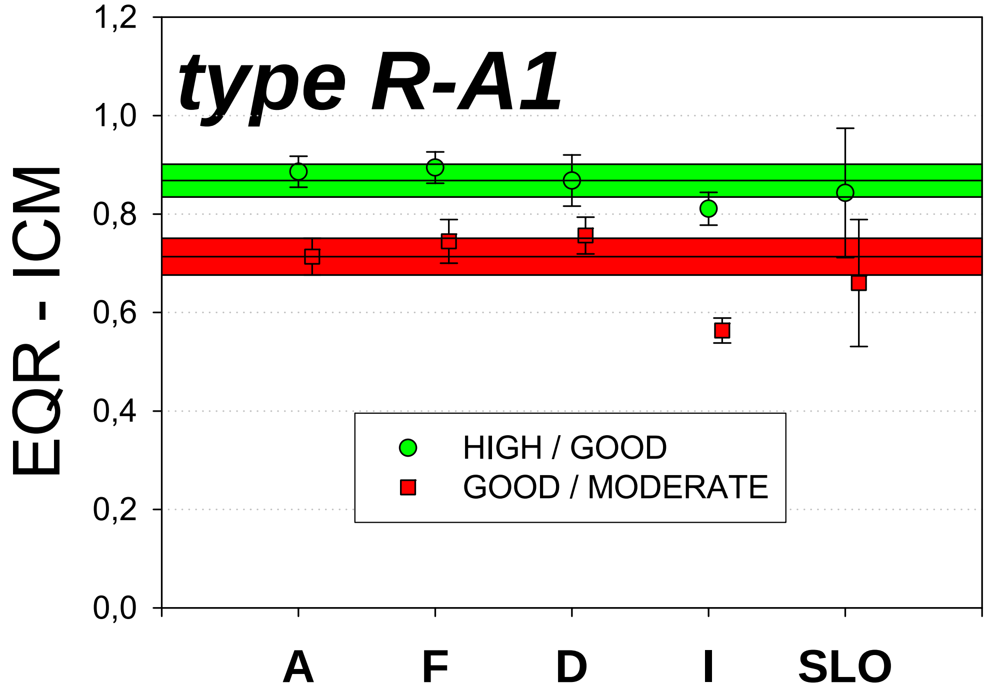


Figure 20: Correlation between the EQR values from the national methods and the EQR values from the single metrics as well as the ICMi *qualitative* and the ICMi *quantitative*.



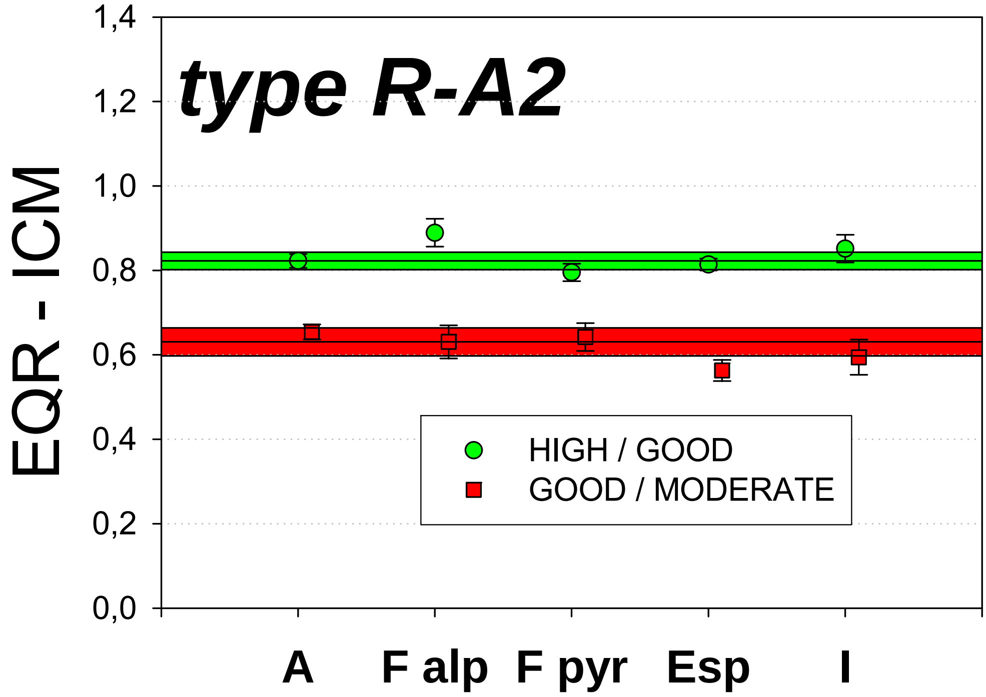


Figure 21: Boundary values of the member states for the Alpine GIG types as calculated from the regression between national method and ICMi *qualitative* values. Given are the values for the high/good and the good/moderate boundary and the 95% confident limits, expressed in EQR of the ICMi method. The bands represent the median boundary values of all member states **+/- the median confidence limits** of all member states.

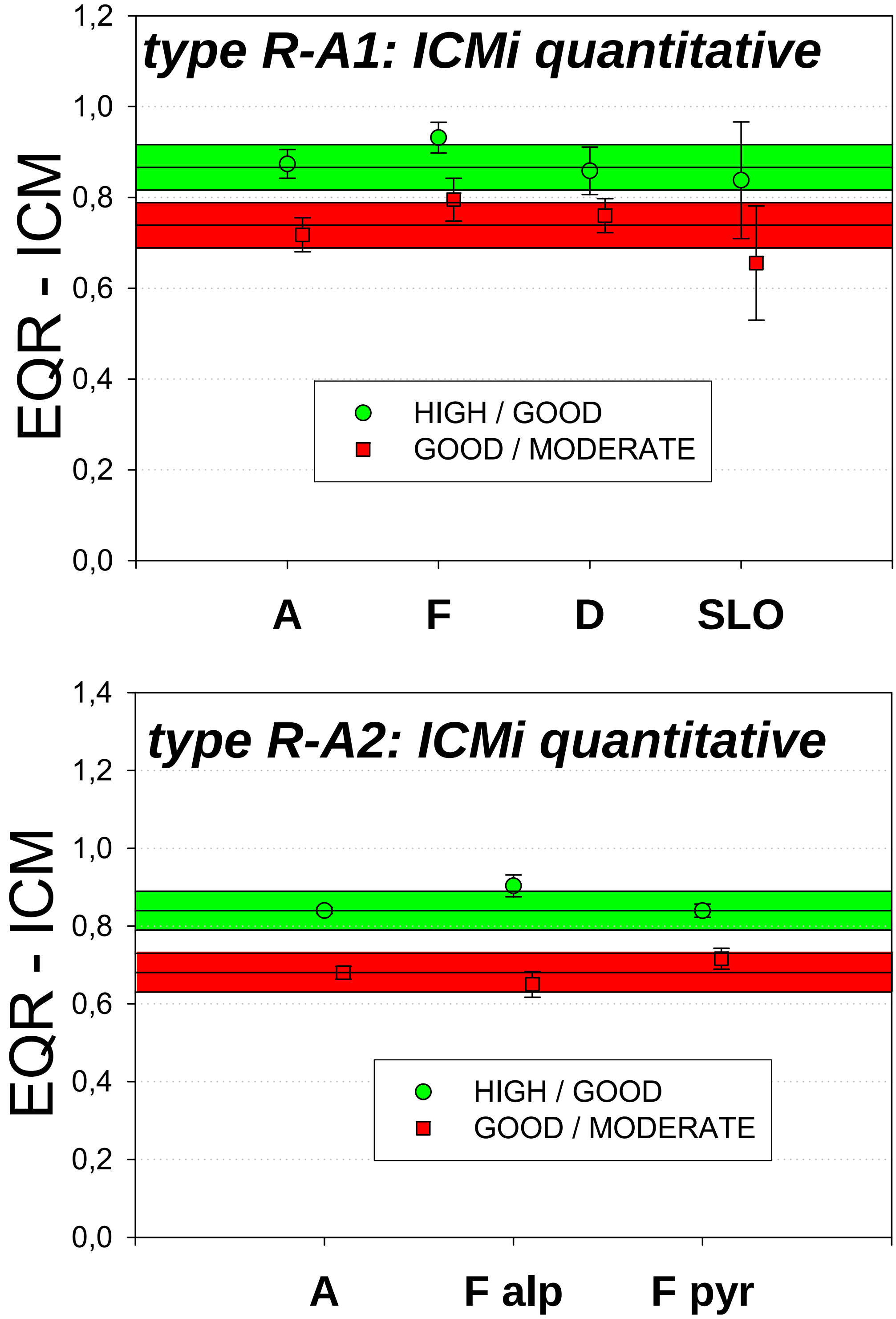


Figure 22: Boundary values of the member states for the Alpine GIG types as calculated from the regression between national method and **ICMi *quantitative*** values. Given are the values for the high/good and the good/moderate boundary and the 95% confident limits, expressed in EQR of the ICMi method. The bands represent the median boundary values of all member states +/- ¼ of the median good status class range of all member states.