Assessing the performance of land-use models: concepts and spatial validation

Sept. 21, 2006

Definition: Calibration

- "Calibration refers to the process of creating a model such that it is consistent with the data used to create the model" (Verburg et al 2006)
- Derivation of best-fit model parameters from real-world data (Dawn)
- Goodness of fit measures of calibration only measure model applicability within the range of calibration data. (We will see a lot of this for econometric models)

Definitions: Verification and Validation

- “Verification and validation concern, respectively, the correctness of model construction and the truthfulness of a model with respect to its problem domain. In other words, verification means building the system right, and validation means building the right system…Once a model is verified and works correctly, then the modeler is concerned with validation—comparing model outcomes to outside data and expectations.” (Manson from Parker et al. 2003)

Model Verification

- Not as much an issue for this class--most important for deductively-structured models (mathematical programming and agent-based)
- The main idea is that you must test out your model rules methodically to make sure they are doing what you intend them to do
- Verification can also be similar to structural validation--are the rules the right rules?

Model validation: Verburg et al 2006

Rykiel (1996) defines validation as “a demonstration that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model”. Model validation is therefore the process of measuring the agreement between the model prediction and independent data. If there is a good match, then the method used to make the prediction is said to be valid. It is crucial to distinguish between model calibration and model validation.

Goals of validation (Verburg et al 2006; Manson 2003)

- Manson, answer question: “How well does a model characterize the target system?”
- Verburg et al: “It is not particularly useful to attempt to crown a model as valid, or to condemn a model as invalid based on the validation results. It is more useful to state carefully the degree to which a model is valid. Validation should measure the performance of a model in a manner that enables the scientist to know the level of trust that one should put in the model. Useful validation should also give the modeller information necessary to improve the model.”

General validation criteria (Quote Manson 2003 from Kwansnicki 1999)

- Correctness: model structure and outcomes must be similar to those of the target system.
- Consistency: the model must be internally consistent and match the conceptual framework in order to describe the target system.
- Universality: the model should be applicable to circumstances beyond those described by the calibration data.
- Simplicity: when choosing between two models, all other things being equal, the less complicated model is preferable.
- Novelty: a model should create new knowledge or outcomes.

Competing measures?

- Note there may be conflicts between these criteria
- Correctness may be more important for empirical, case specific models
- Novelty may be more important for generalizable, theoretical models
- Manson notes (p. 61) that the final structure of the model may be the outcome of interest

Structural vs. outcome validation

- Structural validation examines whether the mechanisms in your model correctly represent real-world mechanisms: are the exogenous components of your model representative of the system under study?
- Outcome validation asks whether model output (endogenous components) conform with data derived from the real-world system
- Outcome validation can be either spatial or a-spatial; today's articles mainly focus on spatial validation

Important issues in calibration, verification, and validation

- Theory can play an important role in model verification and structural validation
- Extensive sensitivity analysis should be conducted for complex models
- Separation between data used for calibration and data used for validation must be maintained
- Success of validation may be influenced by the scale and resolution of data and analysis
- For a nice overview of V&V issues in LUCC models, see Manson: http://www.csiss.org/maslucc/ABM-LUCC.htm#_Toc24263609

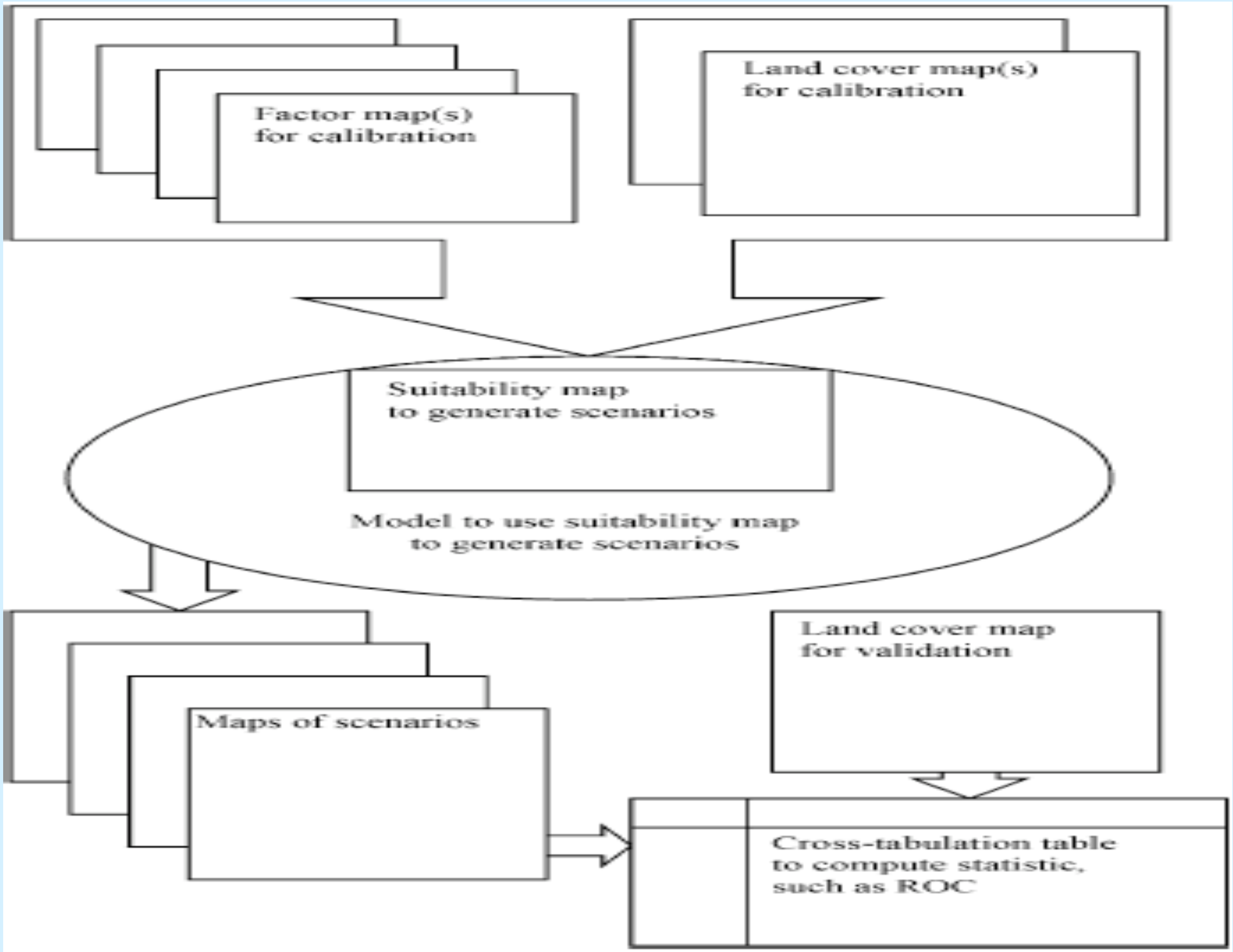
Concepts in spatial validation: map comparisons

- Land-use composition at differing extents
- Absolute or relative location of land use
- Relationship between predicted land use and local spatial characteristics
- Ability of model to correctly predict change
- Spatial pattern: patch size, adjacency, contiguity, fragmentation, clustering, density, nearest neighbor distances, connectivity, etc.
- All may be important

Concepts in spatial validation: Models with many outcomes

- Models may have many possible outcomes due to:
  - Path dependence related to initial conditions
  - Path dependence related to stochastic processes within the model
- Techniques for analyzing multiple outcomes:
  - Generation and testing of distirbutions of outcomes (Lewis and Plantinga)
  - Analysis of variant vs. invariant regions of change (Brown et al)

From Pontius and Schneider 2001



Oreskes et al.: caveats to validation

- NOTE: They use the reverse definition of verification and validation compared to most of the land-use modeling community
- Focuses on models used for predictive purposes and offers some caveats
- Complete V/V is only possible in a closed system, and natural systems are not closed
- Scaling up may introduce error
- Many model formulations can produce the same output (non-identification in statistics, or non-uniqueness)

Oreskes et al. continued

- They don’t like the term “validation”; well, we need to use some term, and need to compare our models to reality, so I disagree.
- They note that on the verification side, the ability of a model to replicate an analytical solution does not imply that it is a representation of the real world; I agree with this.
- Concept of “empirical adequacy” is a useful one, as is concept of “confirmation”

Oreskes et al.: what good are models then?

- Corroborate a hypothesis
- Elucidate discrepancies in other models
- Sensitivity analysis
- "The reason for modeling is a lack of full access, either in time or space, to the phenomena of interest" (p.644)

Turner et al.: Pattern and scale

- Historically (note this article is from 1989) ecological simulations had been spatially aggregated
- Spatial pattern can reflect function such as flow of energy, species, and materials (see Parker and Meretsky for a more recent brief review)
- Therefore, comparisons of predicted and actual spatial patterns are called for

Turner et al.: Measures of spatial pattern

- Fractal dimension:
  - Can be based on perimeter/area relationships and/or patch size distributions.
  - Calculated via regression analysis based on hypothesis that the relationships are linear in logs
  - Good fit for the fractal model implies a power law distribution of edge lengths and/or patch sizes.
  - Higher values of the estimated parameter imply more convoluted and/or fragmented landscapes

Turner, spatial pattern, cont.:

- Adjacency and contagion
- Nearest neighbor probabilities represent proportion of adjacencies
- Contagion measures extent of clumping
- Edge differences are also important
- All three are correlated with each other and with patch size

Turner et al., spatial predictability

- Depends on correlation between location parameters (local characteristics and state of neighboring land uses)
- Closely related to spatial econometric concepts; stay tuned
- Predictability is a measure of spatial structure

Turner et al., multiple resolution goodness of fit

- “A cell by cell comparison between a model’s predicted spatial patterns and the actual patterns is necessary if the location of different habitats relative to their actual location is important” (p. 8)
- Fit at local spatial scale is determined by the proportion of cells correctly predicted, regardless of arrangement
- In addition to reporting averages, comparisons can be mapped out
- At spatial extent of the model, method compares predicted to actual composition

Turner et al., spatial examples

- Provide a very nice set of examples comparing measures across generated landscapes
- Table 7 provides a nice summary of implications
- Spatial pattern measures may be useful for predicting the behavior of a system
- Multiple resolution goodness of fit measures provide info on location matching, but are not sensitive to differences in pattern
- Short lesson is that both may be useful

Pontius, Quant vs. Location

- Paper distinguished between quantification error (error in estimation of proportion in each land use, or composition) and location error (error in predicting location of land uses)
- It also introduces indices designed to measures model success in predicting proportion correct due to change, quantity prediction, and location prediction

Contingency tables based on composition only

TABLE 1. CONTINGENCY TABLE FOR J CATEGORIES WHERE ENTRIES ARE PROPORTIONS OF STUDY AREA

Simulation	Reality				
	1	2	…	J	total
1	p_{11}	p_{12}		p_{1J}	$S_1 = \Sigma p_{1j}$
2	p_{21}	p_{22}		p_{2J}	$S_2 = \Sigma p_{2j}$
…					…
J	p_{J1}	p_{J2}		p_{JJ}	$S_J = \Sigma p_{Jj}$
total	$R_1 = \Sigma p_{j1}$	$R_2 = \Sigma p_{j2}$	…	$R_J = \Sigma p_{jJ}$	1

Evaluating the contingency table

- Proportion correct does not account for effects of being correct by chance
- Kappa statistic is an index that corrects for the proportion correct due to chance:

$$\text{Kappa} = \frac{(Po - Pc)}{(Pp - Pc)}$$

Kappa alternatives

- **Kno** "indicates the proportion classified correctly relative to the proportion classified correctly by a simulation with no ability to specify accurately quantity or location" (p. 1013). Claims to solve problem with Kappa:
- Penalizes for large quantification error
- Rewards for accurate classification
- But does not distinguish between quantity and location error

Pontius, Klocation and Kquantity

- Klocation measures success in predicting location for a model with no ability to predict quantity
- Kquantity measures success in predicting quantity given fixed location

Application: Land-use change in the Ipswich watershed

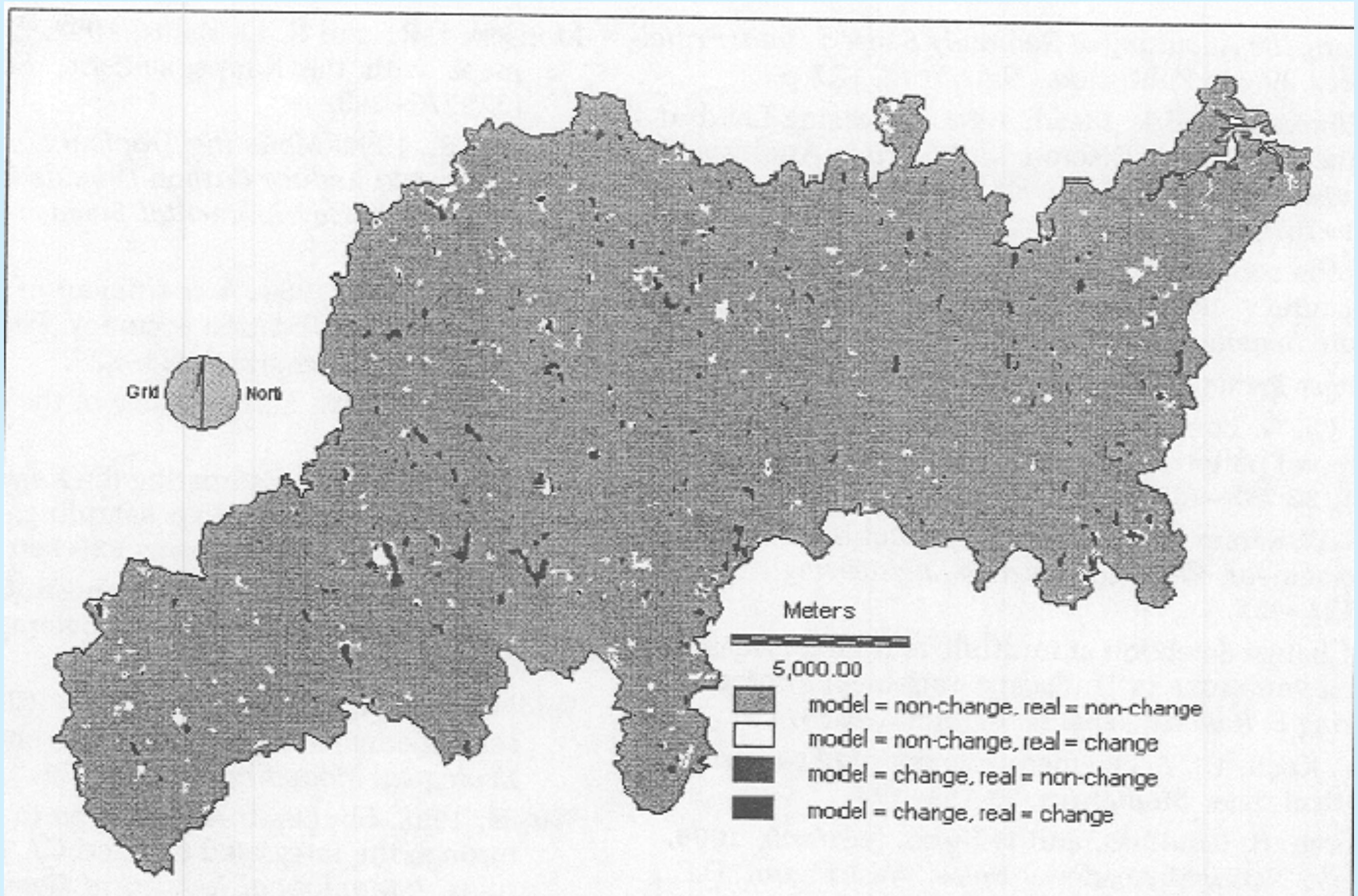


Figure 5. Simulated versus real change to new residential land use from 1985 to 1991 in the Ipswich watershed.

Results, Ipswich study

TABLE 3. FORMULAS AND SIMULATION RESULTS FOR VARIATIONS OF KAPPA

Kappa Definition		Ipswich Results	
Variations	Formula	Formula	Kappa (%)
Kstandard	$\frac{Po - \text{MQNL}}{1 - \text{MONL}}$	$\frac{0.91 - 0.90}{1.00 - 0.90}$	12
Kno	$\frac{Po - \text{NQNL}}{1 - \text{NQNL}}$	$\frac{0.91 - 0.50}{1.00 - 0.50}$	82
Klocation	$\frac{Po - \text{MQNL}}{\text{MQPL} - \text{MQNL}}$	$\frac{0.91 - 0.90}{0.94 - 0.90}$	28
Kquantity	$\frac{Po - \text{NQML}}{\text{PQML} - \text{NQML}}$	$\frac{0.91 - 0.51}{0.97 - 0.51}$	87

Evaluating model results

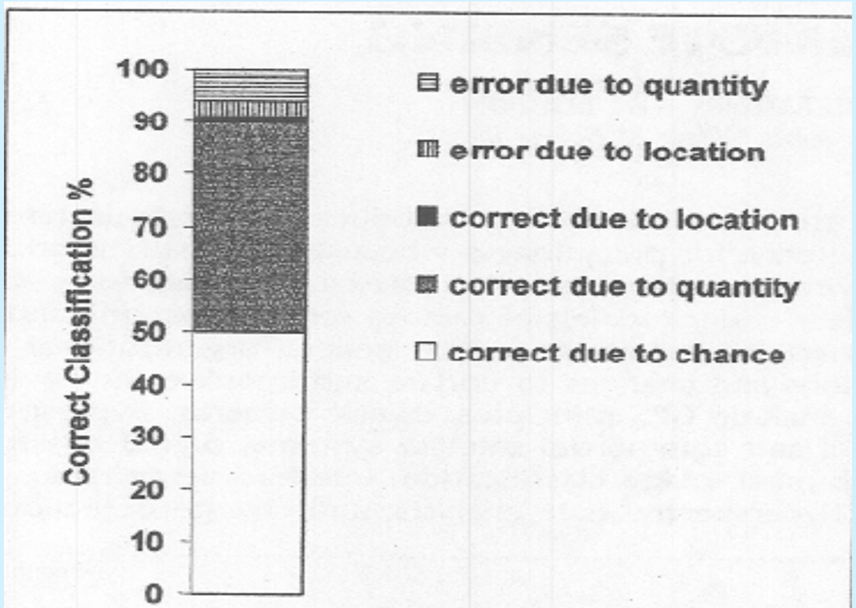


Figure 6. Sources of classification successes and errors assuming fixed specification of quantity in the Ipswich watershed.

- Model's success is due to its ability to predict quantity correctly
- Model does fairly poorly at location
- Much of success is attributable to chance
- Note model predicted more change (8%) than really occurred (2%)
- Note location and quantity will change together

Pontius and Schneider

- In the previous paper, we saw that the region under study was characterized by large areas of persistence of land use, and the model poorly predicted change
- This paper focuses on the concept of validation by comparing areas of predicted and actual change and non-change

Contingency tables for change and no change

Table 1
Two-by-two contingency table showing the proportion (or number) of grid cells in a map of reality versus a map of a modeled scenario

Model	Reality		
	Change	Non-change	Total
Change	A	B	A + B
Non-change	C	D	C + D
Total	A + C	B + D	A + B + C + D

- ROC works for suitability maps where there are two possible land-use types
- Appropriate for models that produce multiple simulated realizations, or probabilistic predictions
- Ranks cells by suitability in percentiles
- Suitability maps are overlaid with maps of reality

- Tables show the proportion of grid cells in each suitability group (assumed to change) classified as change in the real-world map
- Rates of true and false positives are reported
- ROC measures how closely simulated change tracks real change
- Has a value of 0.5 for random assignment, and 1 for perfect assignments

Ipswich example

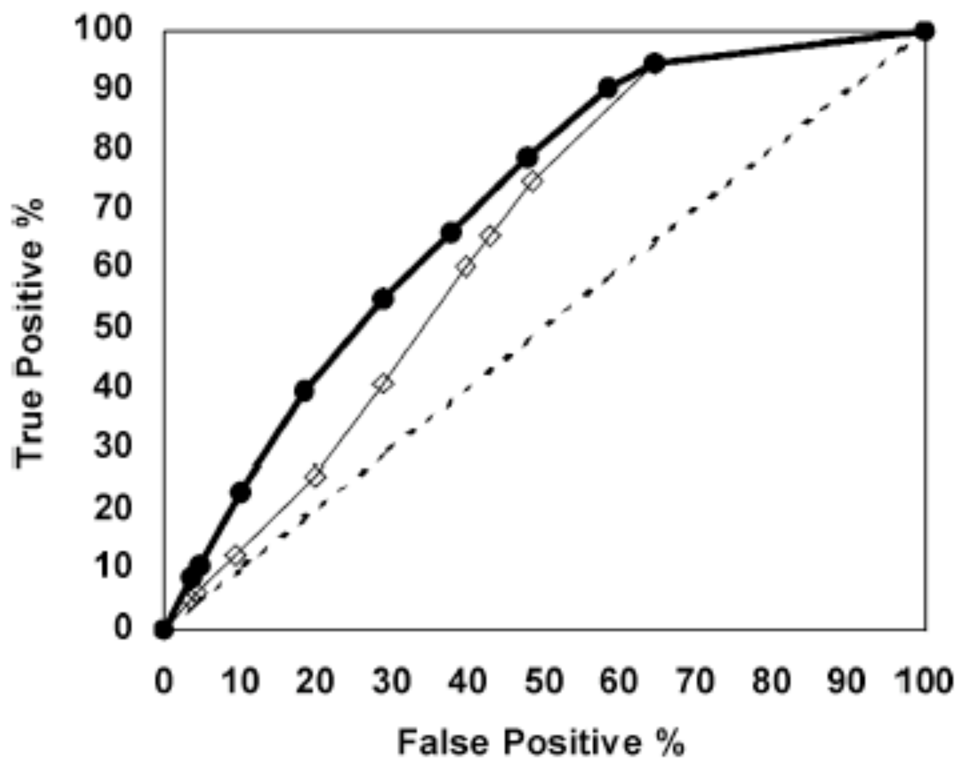


Fig. 5. ROC curves to validate models of deforestation in the Ipswich watershed between 1985 and 1991 using three suitability maps based on: random location (bottom ROC = 50%), logistic regression with protected areas (middle ROC = 65%), and MCE with a spatial filter and protected areas (upper ROC = 70%).

Comments

- ROC does not account for spatial arrangement
- This paper provides a nice head-to-head comparison of different models
- Other work by Pontius has extended previous work to multi-resolution goodness of fit measures
- Recent work by Pontius focuses on the ability of models to predict change using other techniques (Pontius 2004: “Detecting important categorical land changes while accounting for persistence”, AEE 101)