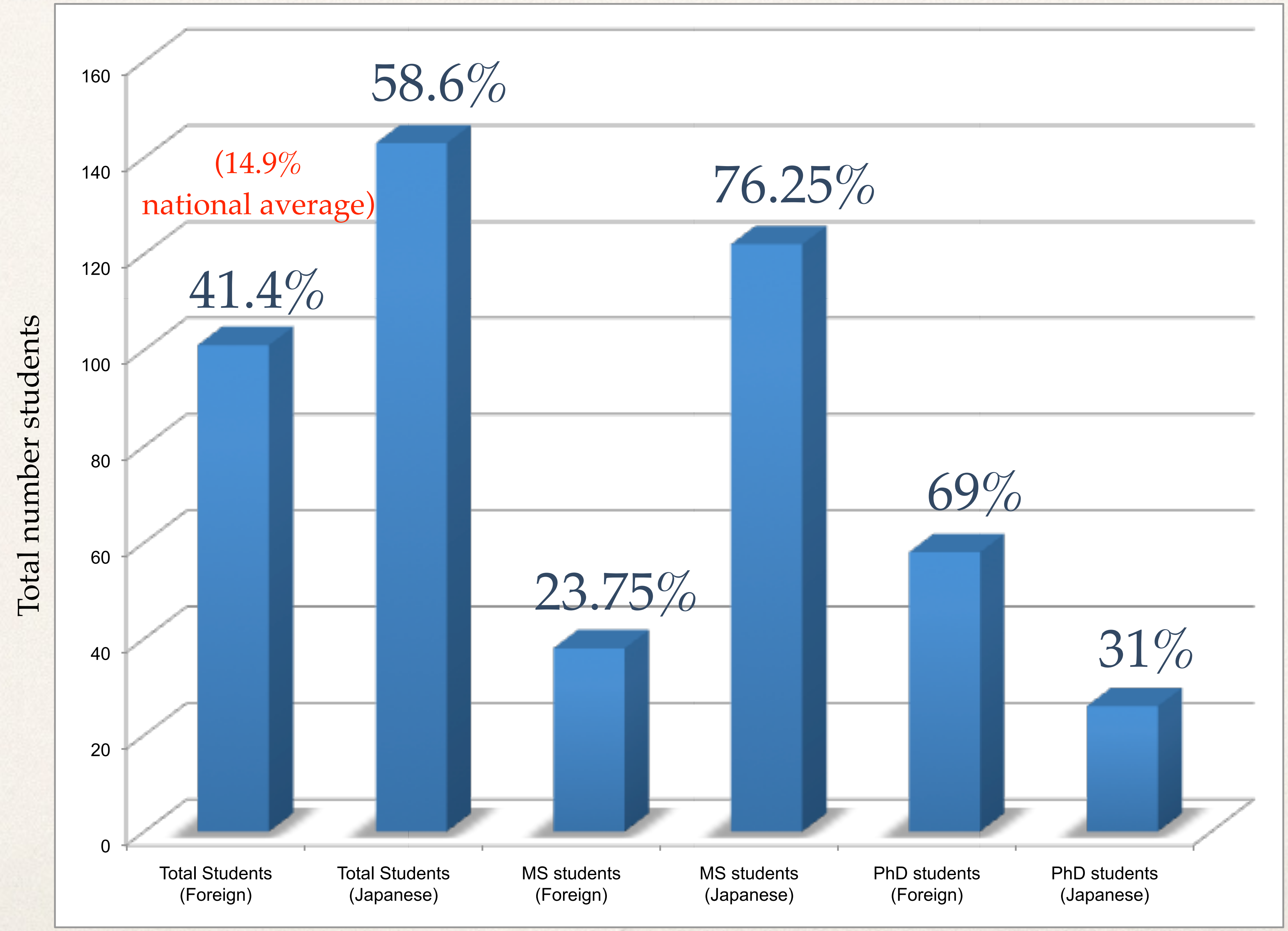
Building a specialised corpus of civil engineering research articles (SCCERA)

Alex Gilmore
Department of Civil Engineering
University of Tokyo, Japan

September 4, 2014

Department of Civil Engineering, University of Tokyo

Postgraduate student population - Department of Civil Engineering, University of Tokyo



Incoming students: Where are they from?

Why create a specialised corpus?

- Large variation between different academic disciplines in terms of word frequencies, collocational patterns & rhetorical moves: e.g. 4-word lexical bundles from fields of Biology, Electrical Engineering, Applied Linguistics & Business Studies >50% unique (Hyland 2008)
- Specialised corpora a good starting point for design of ESP materials for post-graduate students & staff in Department of Civil Engineering

Research questions

- What are the most frequently occurring words, keywords or 3 - 8-word bundles in civil engineering RAs?
- What are the most frequently occurring lexico-grammatical patterns?
- Do any general high-frequency words take on discipline-specific meanings (e.g *wicked* problems)?
- Are any genre-specific move sequences identifiable in abstract, introduction & discussion sections?
- What pedagogically useful patterns are identifiable?

Pedagogically motivated research questions

- To what extent can corpus-informed materials help post-graduate students & staff to write in discipline appropriate ways?
- Can a more direct approach (civil engineers querying SCCERA themselves) be effective?

Research project implemented in 5 phases

- **Phase 1**: Consultation with corpus linguists & civil engineers on the design & make-up of SCCERA (balanced & representative)
- **Phase 2**: Construction of SCCERA
- **Phase 3**: Quantitative & qualitative analysis of the corpus
- **Phase 4**: Exploring pedagogic applications of the corpus
- **Phase 5**: Dissemination of research results

Phase 1: Consultation on corpus design

Corpus linguists & academics from 12 departments of Civil Engineering consulted on design criteria:

- Peer-reviewed journals, preferably listed in Science Citation Index Expanded (SCI) or Social Sciences Citation Index (SSCI)
- Widely read & respected by researchers; considered “key journals” or “desired outlets for academic work”
- Higher impact factors (IF), 5-year IF, Eigenfactor, article influence (Thomson Reuters)
- Research articles selected by (a) Most cited; (b) Most viewed; (c) Most recent (1 article per volume)
- Minimum size of 1 million words recommended for specialised corpora (Kennedy 1998; Pearson 1998; Rea Rizzo 2010)

Department of Civil Engineering

1. Infrastructure Development & Construction Management
2. Landscape Planning & Design
3. Regional Planning & Surveying
4. Transportation Engineering & Planning
5. River & Environmental Engineering
6. Coastal & Ocean Engineering
7. Hydrology & Water Resources Engineering
8. Geotechnical Engineering
9. Concrete & Construction Engineering
10. Earthquake & Disaster Mitigation Engineering
11. Mechanics & Structures
12. International Projects

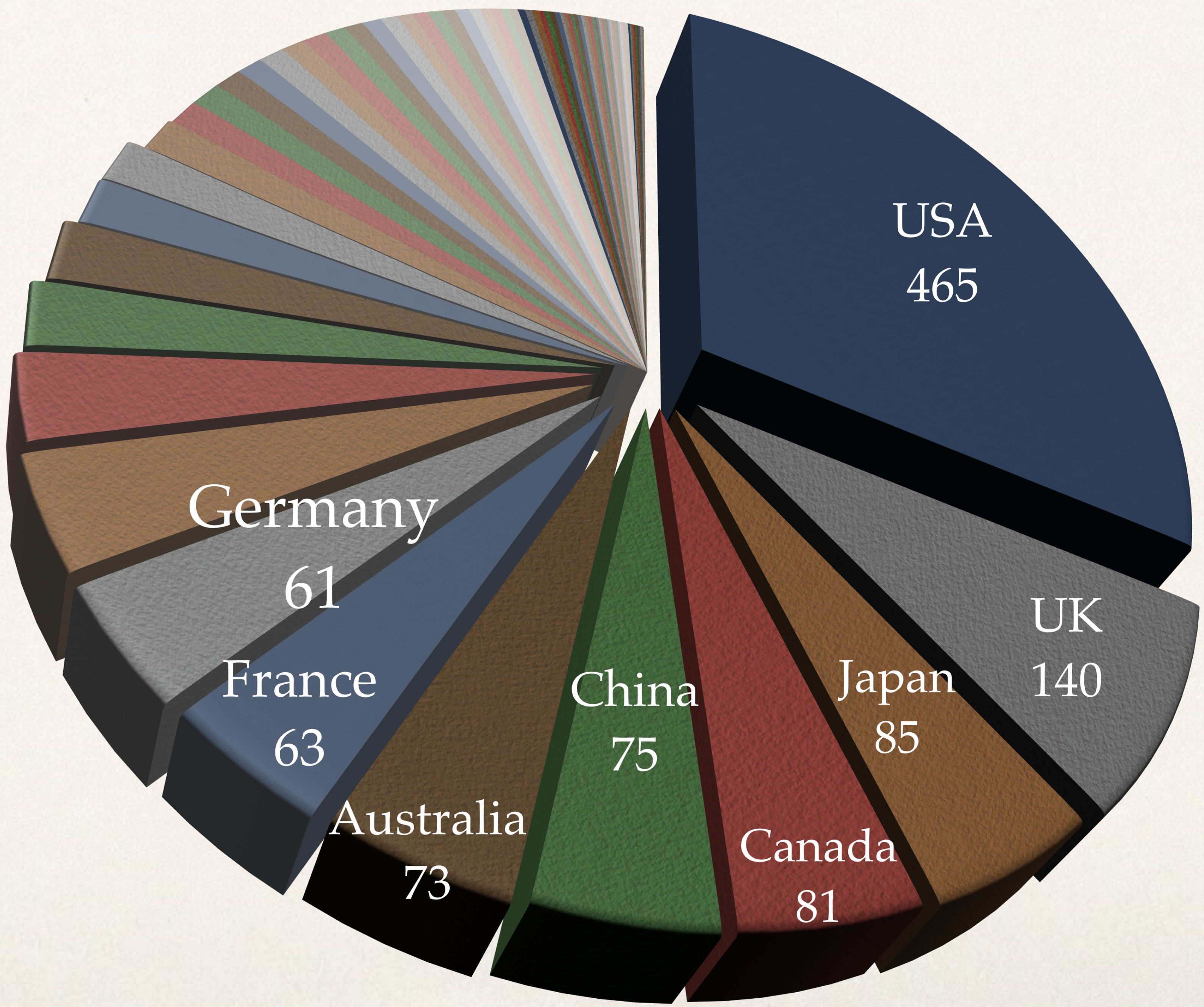
SCCERA journal list

Journal	Article Title	Article Code	Number Authors	Institution Countries	Year of Publication	Number Words
Coastal Engineering	Modelling storm impacts on beaches, dunes	CE_1	6	Netherlands; USA	2009	11,398
Coastal Engineering	Corrected Incompressible SPH method for a	CE_2	3	Japan; UK	2008	8,184
Coastal Engineering	44-year wave hindcast for the North East At	CE_3	3	Portugal; Spain	2008	2,817
Coastal Engineering	Increasing wave heights and extreme value	CE_4	3	USA	2010	10,302
Coastal Engineering	Modified Moving Particle Semi-implicit meth	CE_5	1	Japan	2009	11,452
Coastal Engineering	Numerical analysis of wave overtopping of r	CE_6	4	Spain	2008	7,208
Coastal Engineering	A 44-year high-resolution ocean and atmos	CE_7	5	Spain; France	2008	8,829
Coastal Engineering	Simulation of nonlinear wave run-up with a	CE_8	2	Denmark	2008	7,499
Coastal Engineering	Beach Wizard: Nearshore bathymetry estim	CE_9	6	Netherlands; USA; Chile	2008	8,286
Coastal Engineering	Efficient computation of surf zone waves usi	CE_10	2	Netherlands	2008	6,532
Coastal Engineering	An integrated model for the wave-induced s	CE_11	4	UK; China; USA	2013	8,876
Coastal Engineering	Statistical simulation of wave climate and e	CE_12	4	Australia	2008	8,888
Coastal Engineering	Hindcast of the wave conditions along the w	CE_13	3	Portugal	2008	5,480
Coastal Engineering	Laboratory and numerical studies of wave d	CE_14	3	USA	2009	6,662
Coastal Engineering	A probabilistic methodology to estimate futu	CE_15	3	UK	2008	6,178
Coastal Engineering	Run-up of tsunamis and long waves in term	CE_16	2	Denmark	2008	8,813
Coastal Engineering	Measurement of wave-by-wave bed-levels i	CE_17	3	Australia; UK	2008	1,648
Coastal Engineering	The morphological response of a nearshore	CE_18	4	Netherlands; USA	2008	7,938
Coastal Engineering	Direct bed shear stress measurements in bo	CE_19	4	Australia; UK	2009	9,281
Coastal Engineering	Two-dimensional time dependent hurricane	CE_20	7	Netherlands; USA	2010	9,300
Coastal Engineering	Modeling hurricane waves and storm surge	CE_21	10	Netherlands; USA	2011	9,528
Coastal Engineering	On the evolution and run-up of breaking sol	CE_22	4	Taiwan	2008	7,007
Coastal Engineering	Morphodynamic responses to the deep wate	CE_23	3	China	2009	7,962
Coastal Engineering	Large-scale dune erosion tests to study the	CE_24	5	Netherlands	2008	5,852
Coastal Engineering	Wave boundary layer over a stone-covered	CE_25	4	Denmark	2008	11,155
J. of Coastal Research	The Role of Remote Sensing in Predicting ar	JCR_1	1	USA	2009	7,568
J. of Coastal Research	Shoreline Definition and Detection: A Review	JCR_2	2	Australia	2005	5,895
J. of Coastal Research	Erosion Hazard Vulnerability of US Coastal C	JCR_3	3	USA	2005	5,147
J. of Coastal Research	A Simple Method of Measuring Beach Profile	JCR_4	2	Portugal	2006	2,389
J. of Coastal Research	A New Global Coastal Database for Impact a	JCR_5	9	Greece; UK; Ireland; Germany; N	2008	4,788
J. of Coastal Research	Assessment of Vulnerability and Adaptation	JCR_6	1	Germany	2008	8,688
J. of Coastal Research	Sustainable Management of Surfing Breaks:	JCR_7	4	New Zealand	2009	11,534
J. of Coastal Research	Importance of Coastal Change Variables in I	JCR_8	3	USA	2010	4,311
J. of Coastal Research	The Healing Sea: A Sustainable Coastal Oce	JCR_9	2	Belgium	2009	11,155
J. of Coastal Research	Open-Ocean Barrier Islands: Global Influenc	JCR_10	2	USA	2011	7,538
J. of Coastal Research	Tracking Oil Slicks and Predicting their Traje	JCR_11	1	USA	2010	6,583
J. of Coastal Research	Classification of Coasts	JCR_12	1	USA	2004	8,476
J. of Coastal Research	Coastal Classification: Systematic Approach	JCR_13	1	USA	2004	20,515

SCCERA characteristics

- Total size: ~ 8 million words
- 45 journals (43 cited in SCI Expanded or SSCI)
- 1,100 research articles (average of 7,324 words per article)
- Year of publication: Range = 1989 - 2014; Mean = 2009
- 3,807 contributing authors (average of 3.46 authors per article)
- 1,598 participating institutions from 80 countries

Participating institutions by country (N = 80)



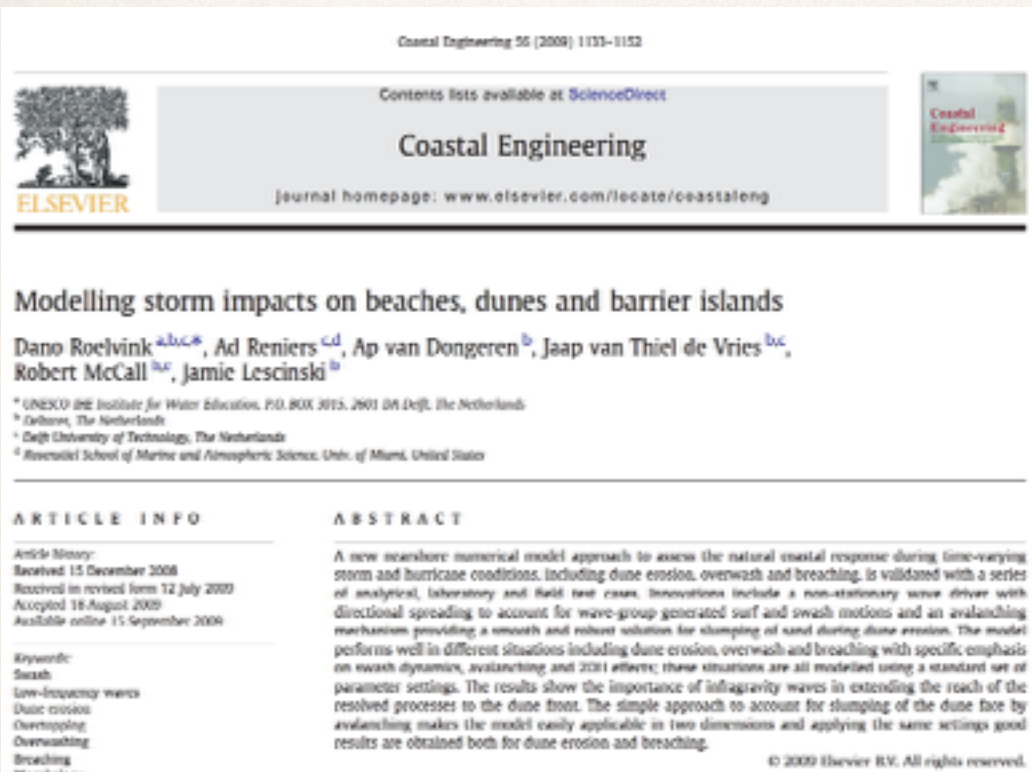
Phase 2: Construction of SCCERA

- HTML or PDF version of articles copied into MS Word
- Extraneous information removed (references, date of acceptance, author affiliation, contact info., tables & figures, equations)
- Text cleaned up using spelling & grammar checking function of MS Word (hyphenated words, conjoined words, character misreadings)
- HTML fragments (‘Table options’, ‘Turn Mathjax on’, etc.) removed using find & replace function in MS Word
- Articles saved as text-only (.txt) files

Phase 2: Construction of SCCERA

- 2nd round of cleaning up using text-only files (Greek symbols, etc.)
- Final document checked against original PDF file
- SCCERA part-of-speech (POS) tagged using CLAWS 4 (Lancaster University UCREL C7 tag set (Total no. tag types = 137): http://ucrel.lancs.ac.uk/claws7tags.html)

Phase 2: Construction of SCCERA


Coastal Engineering 56 (2009) 1133–1152

Contents lists available at ScienceDirect

Coastal Engineering

journal homepage: www.elsevier.com/locate/coastaleng

Modelling storm impacts on beaches, dunes and barrier islands

Dano Roelvink [a,b,c,*], Ad Reniers [c,d], Ap van Dongeren [b], Jaap van Thiel de Vries [b,c], Robert McCall [b,c], Jamie Lescinski [b]

[a] UNESCO-IHE Institute for Water Education, P.O. BOX 3015, 2601 DA Delft, The Netherlands
[b] Deltares, The Netherlands
[c] Delft University of Technology, The Netherlands
[d] Rosenstiel School of Marine and Atmospheric Science, Univ. of Miami, United States

ARTICLE INFO

Article history:
Received 15 December 2008
Received in revised form 12 July 2009
Accepted 18 August 2009
Available online 15 September 2009

Keywords:
Swash
Low-frequency waves
Dune erosion
Overtopping
Overwashing
Breaching
Morphology

ABSTRACT

A new nearshore numerical model approach to assess the natural coastal response during time-varying storm and hurricane conditions, including dune erosion, overwash and breaching, is validated with a series of analytical, laboratory and field test cases. Innovations include a non-stationary wave driver with directional spreading to account for wave-group generated surf and swash motions and an avalanching mechanism providing a smooth and robust solution for slumping of sand during dune erosion. The model performs well in different situations including dune erosion, overwash and breaching with specific emphasis on swash dynamics, avalanching and 2DH effects; these situations are all modelled using a standard set of parameter settings. The results show the importance of infragravity waves in extending the reach of the resolved processes to the dune front. The simple approach to account for slumping of the dune face by avalanching makes the model easily applicable in two dimensions and applying the same settings good results are obtained both for dune erosion and breaching.

^ <text>_NULL

^ Modelling_VVG storm_NN1 impacts_NN2 on_II beaches_NN2 ,_, dunes_NN2 and_CC barrier_NN1 islands_NN2 Abstract_VV0@ A_ZZ1 new_JJ nearshore_NN1 numerical_JJ model_NN1 approach_NN1 to_TO assess_VVI the_AT natural_JJ coastal_JJ response_NN1 during_II time-varying_JJ storm_NN1 and_CC hurricane_NN1 conditions_NN2 ,_, including_II dune_NN1 erosion_NN1 ,_, overwash_NN1 and_CC breaching_VVG ,_, is_VBZ validated_VVN@ with_IW a_AT1 series_NN of_IO analytical_JJ ,_, laboratory_NN1 and_CC field_NN1 test_NN1 cases_NN2 ._.

^ Innovations_NN2 include_VV0 a_AT1 non-stationary_JJ wave_NN1 driver_NN1 with_IW directional_JJ spreading_NN1 to_TO account_VVI for_IF wave-group_JJ generated_JJ@ surf_NN1 and_CC swash_VVI motions_NN2 and_CC an_AT1 avalanching_JJ@ mechanism_NN1 providing_VVG a_AT1 smooth_JJ and_CC robust_JJ solution_NN1 for_IF slumping_VVG of_IO sand_NN1 during_II dune_NN1 erosion_NN1 ._.

^ The_AT model_NN1 performs_VVZ well_RR in_II different_JJ situations_NN2 including_II dune_NN1 erosion_NN1 ,_, overwash_NN1 and_CC breaching_VVG with_IW specific_JJ emphasis_NN1 on_II swash_NN1 dynamics_NN ,_, avalanching_VVG and_CC 2DH_FO effects_NN2 ;_; these_DD2 situations_NN2 are_VBR all_DB modelled_VVN using_VVG a_AT1 standard_JJ set_NN1 of_IO parameter_NN1 settings_NN2 ._.

^ The_AT results_NN2 show_VV0 the_AT importance_NN1 of_IO infragravity_NN1 waves_NN2 in_II extending_VVG the_AT reach_NN1 of_IO the_AT resolved_JJ@ processes_NN2 to_II the_AT dune_NN1 front_NN1 ._.

^ The_AT simple_JJ approach_NN1 to_TO account_VVI for_IF slumping_VVG of_IO the_AT dune_NN1 face_NN1 by_II avalanching_VVG makes_VVZ the_AT model_NN1 easily_RR applicable_JJ in_II two_MC dimensions_NN2 and_CC applying_VVG the_AT same_DA settings_NN2 good_JJ results_NN2 are_VBR obtained_VVN both_RR for_IF dune_NN1 erosion_NN1 and_CC breaching_VVG ._.

UCREL CLAWS 7 Tagset

VVG = -ing participle of lexical verb

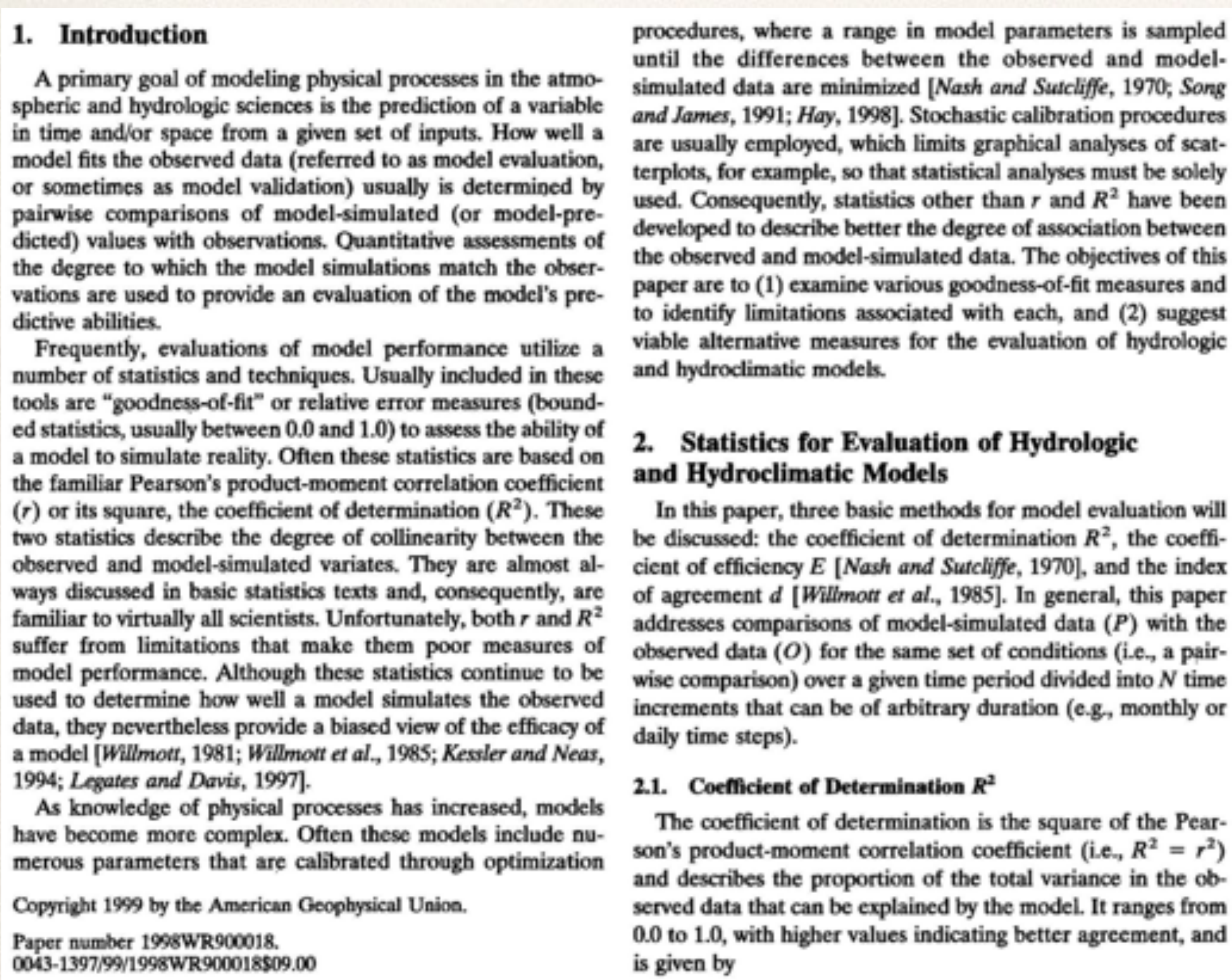
NN1 = singular common noun

NN2 = plural common noun

II = general preposition

Processing time (mins per RA)

Common problems



1. Introduction

A primary goal of modeling physical processes in the atmospheric and hydrologic sciences is the prediction of a variable in time and/or space from a given set of inputs. How well a model fits the observed data (referred to as model evaluation, or sometimes as model validation) usually is determined by pairwise comparisons of model-simulated (or model-predicted) values with observations. Quantitative assessments of the degree to which the model simulations match the observations are used to provide an evaluation of the model's predictive abilities.

Frequently, evaluations of model performance utilize a number of statistics and techniques. Usually included in these tools are "goodness-of-fit" or relative error measures (bounded statistics, usually between 0.0 and 1.0) to assess the ability of a model to simulate reality. Often these statistics are based on the familiar Pearson's product-moment correlation coefficient (r) or its square, the coefficient of determination (R^2). These two statistics describe the degree of collinearity between the observed and model-simulated variates. They are almost always discussed in basic statistics texts and, consequently, are familiar to virtually all scientists. Unfortunately, both r and R^2 suffer from limitations that make them poor measures of model performance. Although these statistics continue to be used to determine how well a model simulates the observed data, they nevertheless provide a biased view of the efficacy of a model [*Willmott*, 1981; *Willmott et al.*, 1985; *Kessler and Neas*, 1994; *Legates and Davis*, 1997].

As knowledge of physical processes has increased, models have become more complex. Often these models include numerous parameters that are calibrated through optimization procedures, where a range in model parameters is sampled until the differences between the observed and model-simulated data are minimized [*Nash and Sutcliffe*, 1970; *Song and James*, 1991; *Hay*, 1998]. Stochastic calibration procedures are usually employed, which limits graphical analyses of scatterplots, for example, so that statistical analyses must be solely used. Consequently, statistics other than r and R^2 have been developed to describe better the degree of association between the observed and model-simulated data. The objectives of this paper are to (1) examine various goodness-of-fit measures and to identify limitations associated with each, and (2) suggest viable alternative measures for the evaluation of hydrologic and hydroclimatic models.

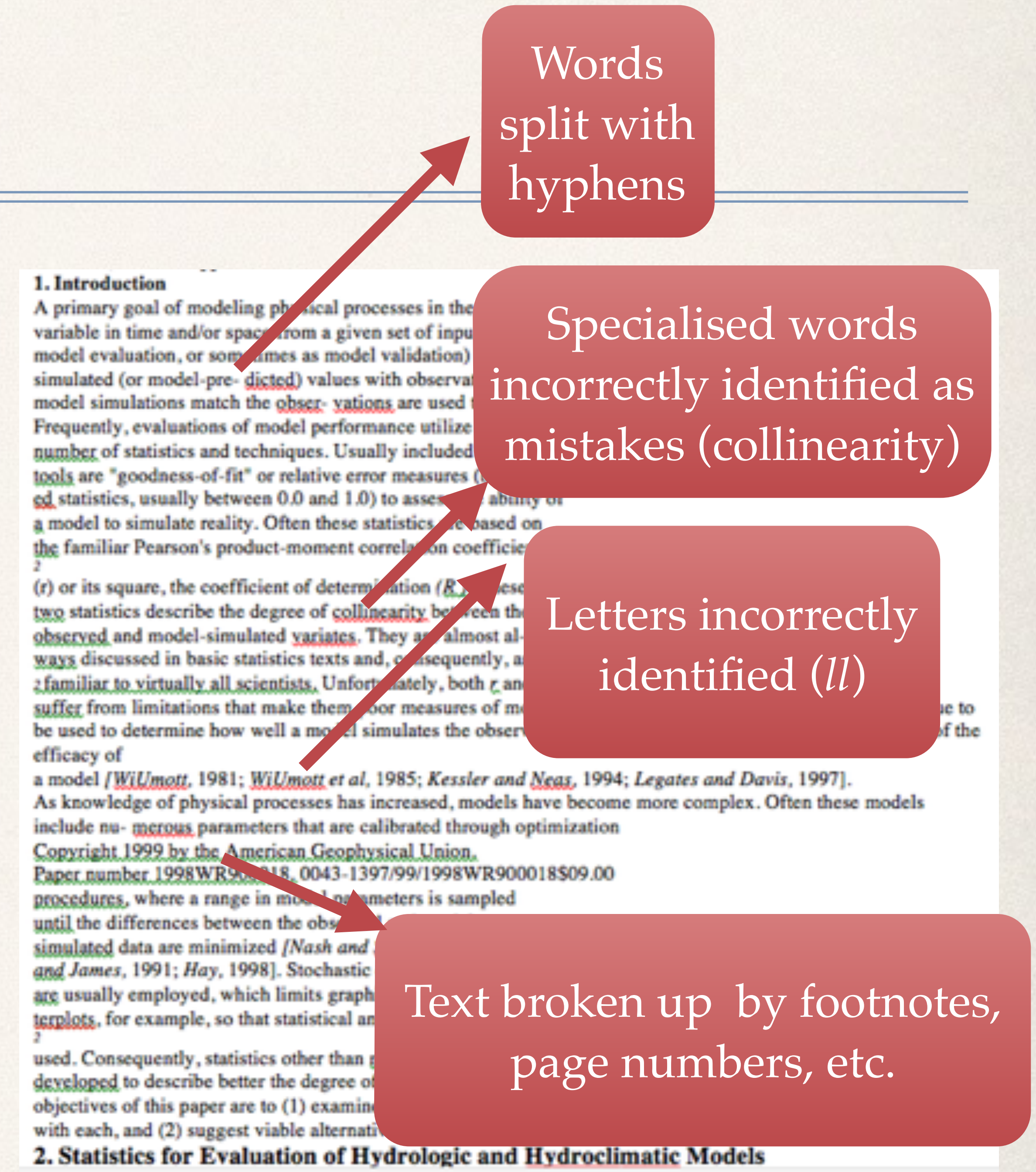
Paper number 1998WR900018.
0043-1397/99/1998WR900018$09.00

2. Statistics for Evaluation of Hydrologic and Hydroclimatic Models

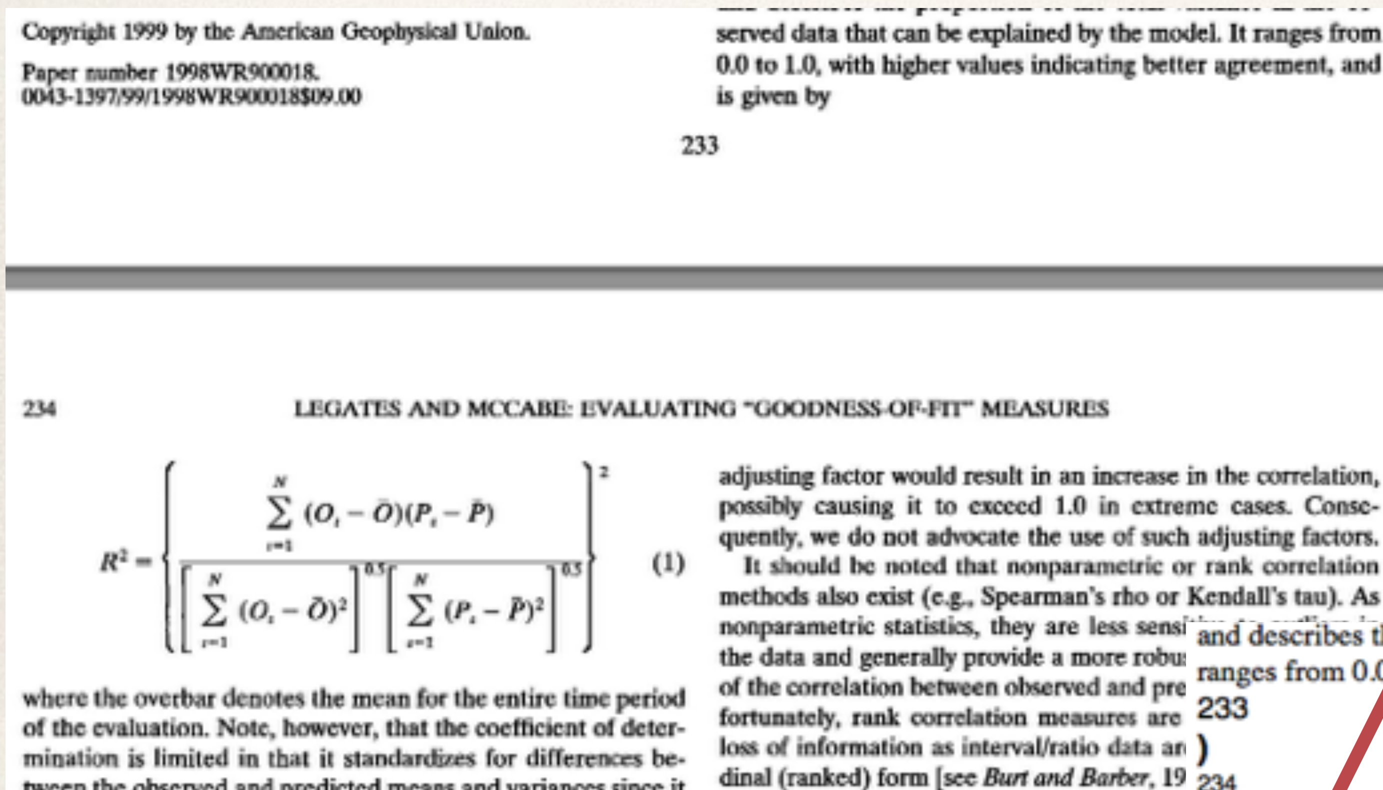
In this paper, three basic methods for model evaluation will be discussed: the coefficient of determination R^2, the coefficient of efficiency E [*Nash and Sutcliffe*, 1970], and the index of agreement d [*Willmott et al.*, 1985]. In general, this paper addresses comparisons of model-simulated data (P) with the observed data (O) for the same set of conditions (i.e., a pairwise comparison) over a given time period divided into N time increments that can be of arbitrary duration (e.g., monthly or daily time steps).

2.1. Coefficient of Determination R^2

The coefficient of determination is the square of the Pearson's product-moment correlation coefficient (i.e., $R^2 = r^2$) and describes the proportion of the total variance in the observed data that can be explained by the model. It ranges from 0.0 to 1.0, with higher values indicating better agreement, and is given by



Common problems



Copyright 1999 by the American Geophysical Union.

Paper number 1998WR900018.
0043-1397/99/1998WR900018$09.00

served data that can be explained by the model. It ranges from 0.0 to 1.0, with higher values indicating better agreement, and is given by

233

234 LEGATES AND MCCABE: EVALUATING "GOODNESS-OF-FIT" MEASURES

$$R^2 = \left\{ \frac{\sum_{i=1}^{N} (O_i - \bar{O})(P_i - \bar{P})}{\left[\sum_{i=1}^{N} (O_i - \bar{O})^2\right]^{0.5} \left[\sum_{i=1}^{N} (P_i - \bar{P})^2\right]^{0.5}} \right\}^2 \quad (1)$$

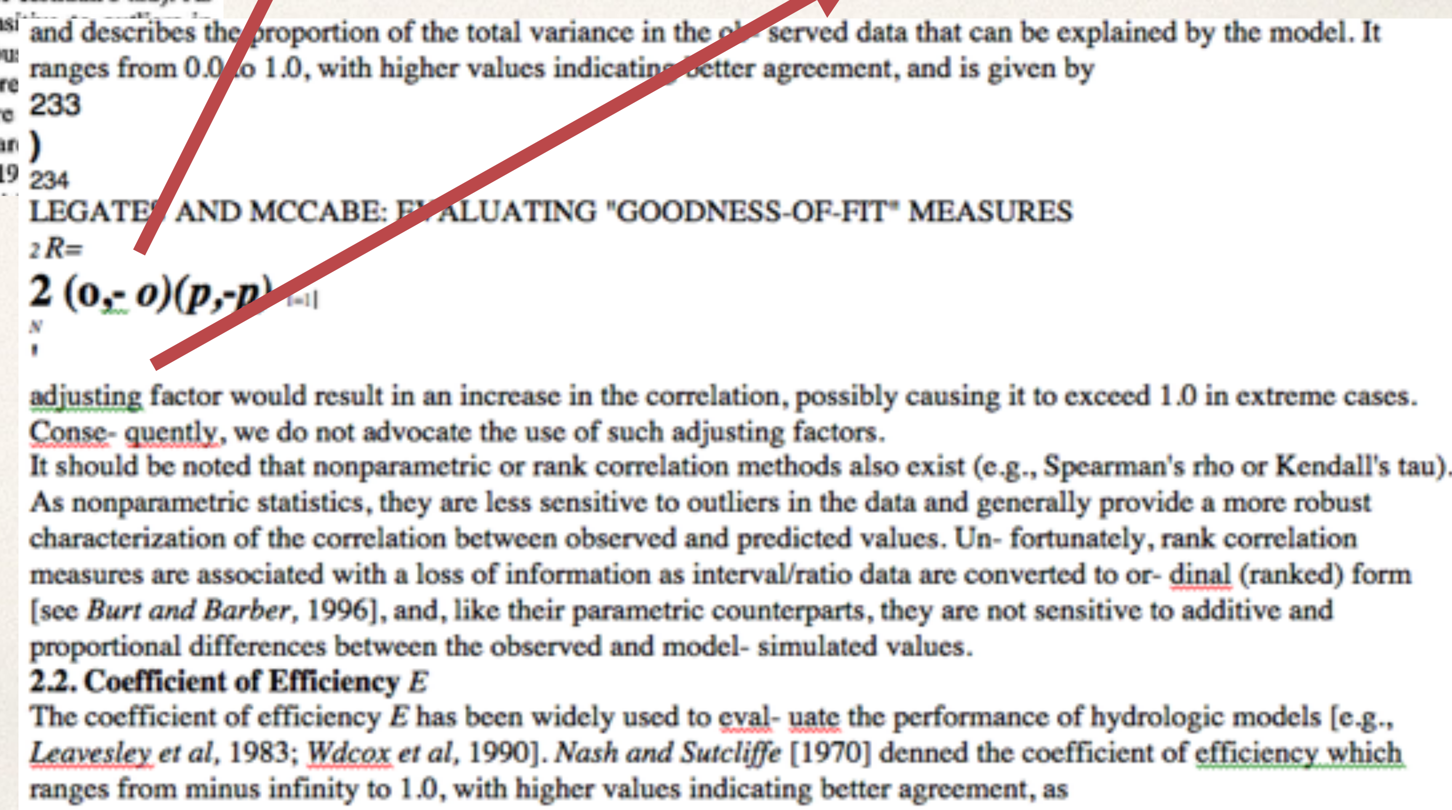
where the overbar denotes the mean for the entire time period of the evaluation. Note, however, that the coefficient of determination is limited in that it standardizes for differences between the observed and predicted means and variances since it

adjusting factor would result in an increase in the correlation, possibly causing it to exceed 1.0 in extreme cases. Consequently, we do not advocate the use of such adjusting factors.

It should be noted that nonparametric or rank correlation methods also exist (e.g., Spearman's rho or Kendall's tau). As nonparametric statistics, they are less sensi[illegible] the data and generally provide a more robu[illegible] of the correlation between observed and pre[illegible] fortunately, rank correlation measures are [illegible] loss of information as interval/ratio data ar[illegible] dinal (ranked) form [see *Burt and Barber*, 19[illegible]

Mathematical symbols not recognised

Columns not recognised

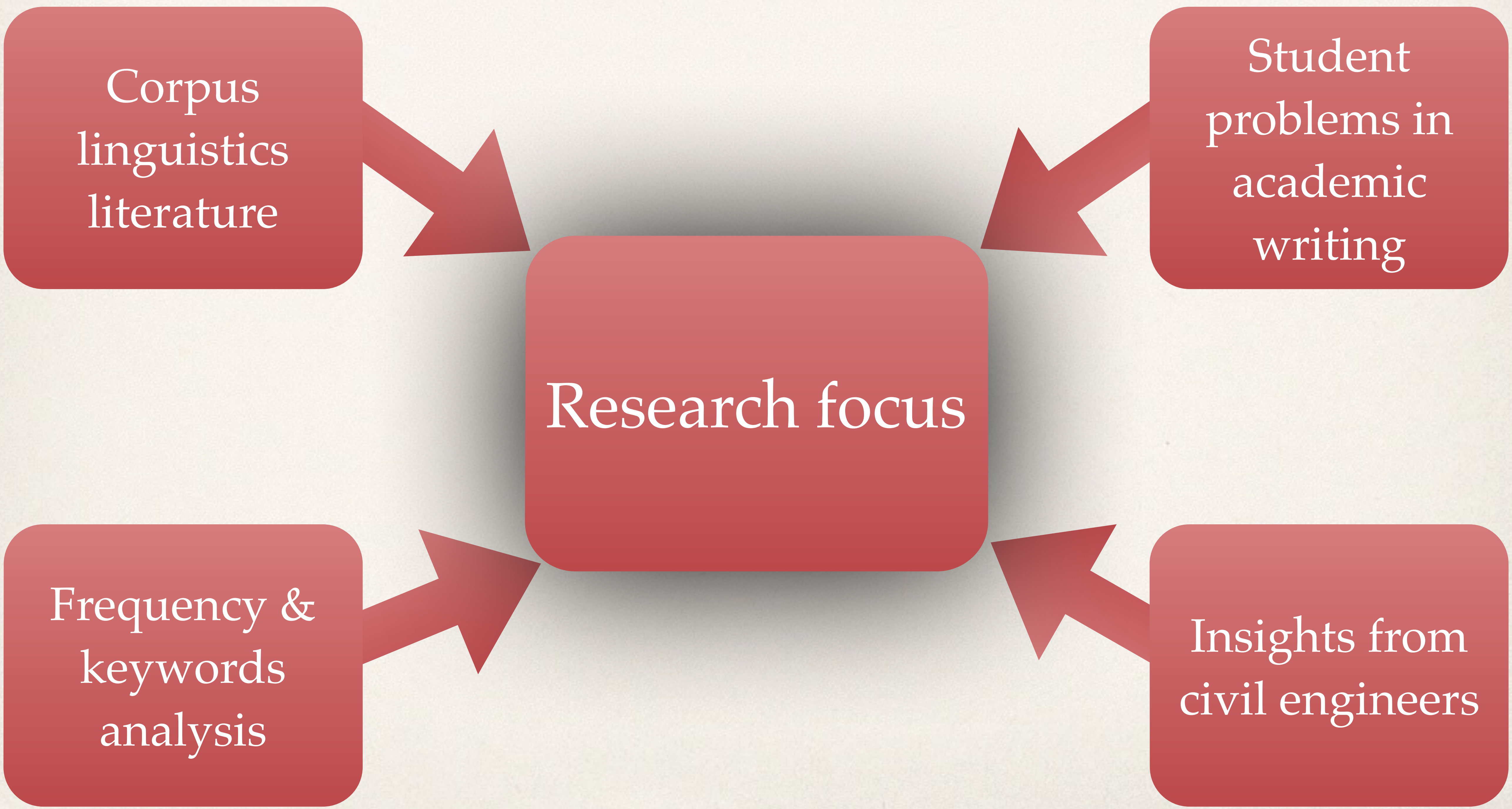


and describes the proportion of the total variance in the ob- served data that can be explained by the model. It ranges from 0.0 to 1.0, with higher values indicating better agreement, and is given by
233
)
234
LEGATES AND MCCABE: EVALUATING "GOODNESS-OF-FIT" MEASURES
2 *R*=
2 (o,- *o*)(*p*,-*p*) i=1]
N
'

adjusting factor would result in an increase in the correlation, possibly causing it to exceed 1.0 in extreme cases. Conse- quently, we do not advocate the use of such adjusting factors.
It should be noted that nonparametric or rank correlation methods also exist (e.g., Spearman's rho or Kendall's tau). As nonparametric statistics, they are less sensitive to outliers in the data and generally provide a more robust characterization of the correlation between observed and predicted values. Un- fortunately, rank correlation measures are associated with a loss of information as interval/ratio data are converted to or- dinal (ranked) form [see *Burt and Barber*, 1996], and, like their parametric counterparts, they are not sensitive to additive and proportional differences between the observed and model- simulated values.
2.2. Coefficient of Efficiency *E*
The coefficient of efficiency *E* has been widely used to eval- uate the performance of hydrologic models [e.g., *Leavesley et al*, 1983; *Wdcox et al*, 1990]. *Nash and Sutcliffe* [1970] denned the coefficient of efficiency which ranges from minus infinity to 1.0, with higher values indicating better agreement, as

[…] and is given by (Equation 1) where x denotes…

Phase 3: Analysis of the corpus



Phase 3: Quantitative analysis of SCCERA

- Corpus analysis using WordSmith Tools 6.0 (Scott 2011)
- Comparisons across (a) RAs, (b) sub-sections, (c) sub-disciplines
- Word frequencies, keywords, key keywords, 2 to 8-word lexical bundles, type/token ratios, pedagogically significant concordance lines - e.g. disambiguation of near-synonymous words (Lee & Swales 2006)

Phase 3: Qualitative analysis of SCCERA

- Discourse analytical approach, investigating rhetorical characteristics of civil engineering RAs
- Move sequences in RA abstracts, introduction & discussion sections (often the most complex & problematic sections)
- Multimodality in civil engineering RAs

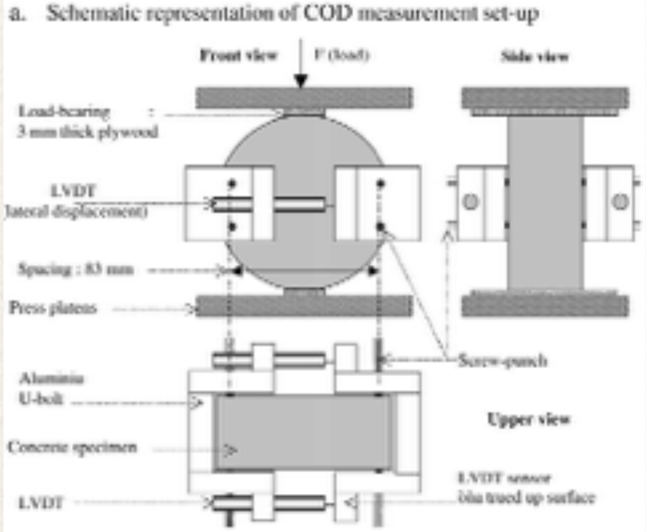
Word frequency (position)

et. al	table	models	concrete (108)
model	figure	flow (93)	method (109)
fig	may	shown	effects (112)
we	values (64)	if (94)	mean (113)
between (36)	level (69)	case (95)	average (114)
time (37)	analysis (72)	large (97)	same (115)
used (39)	surface (76)	project (98)	stress (116)
results (44)	number (77)	area (100)	observed
equation	study (82)	effect (102)	change (126)
using (46)	value (83)	due (104)	see

Multimodality in civil engineering RAs

(see Fig. 1a)

a. Schematic representation of COD measurement set-up



(see Fig. 1)

Photography of specimen during the test

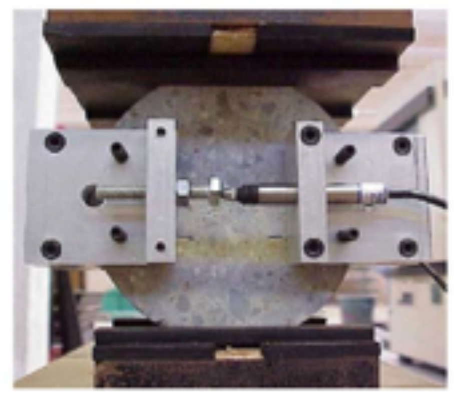


Fig. 1. Controlled splitting test.

[…], as in Eqn. (1):

$$J(x) = -D_e \frac{\partial c}{\partial x} + D_e \frac{zFE}{RTL} c + cv(x) \quad (1)$$

Fig. 3 presents…

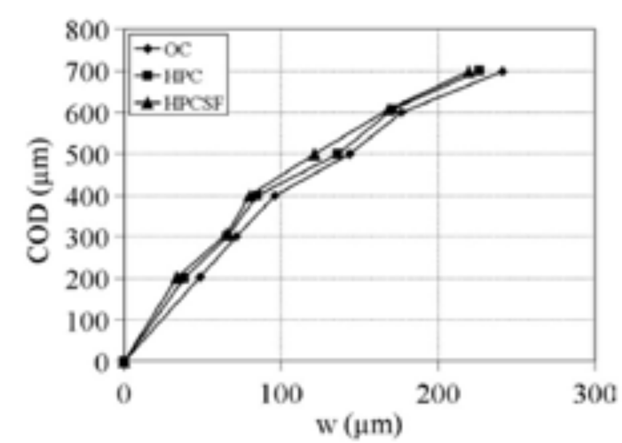


Fig. 3. Crack width versus crack opening displacement under loading.

(see Table 1)

Table 1
Details of test series and mix proportion

Mix ingredients (kg/m3)	OC	HPC	HPCSF
Coarse aggregate, 12.5–20 mm	777	550	579
Medium agrregate, 4–12.5 mm	415	475	465
Sand (Boulonnais), 0–5 mm	372	407	442
Sand (Seine), 0–4 mm	372	401	435
Cement CPA-CEM I 52.5	353	461	360
Silica fume	–	0	22
SP (e.s.)	–	12.4	12
Retarder (e.s.)	–	3.3	2.5
Total water	172	146	136
w/c	0.49	0.32	0.38
w/(C+SF)	0.49	0.32	0.36

(Fig. 2)

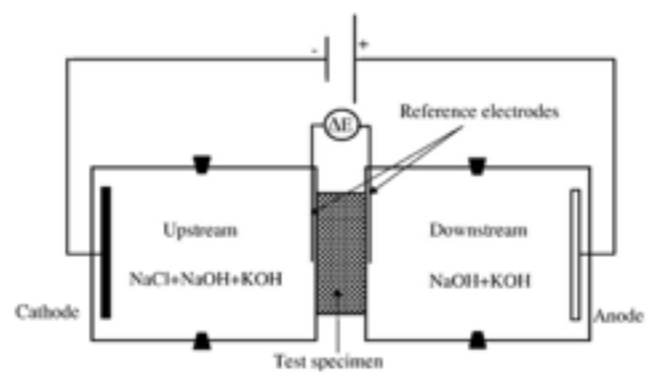


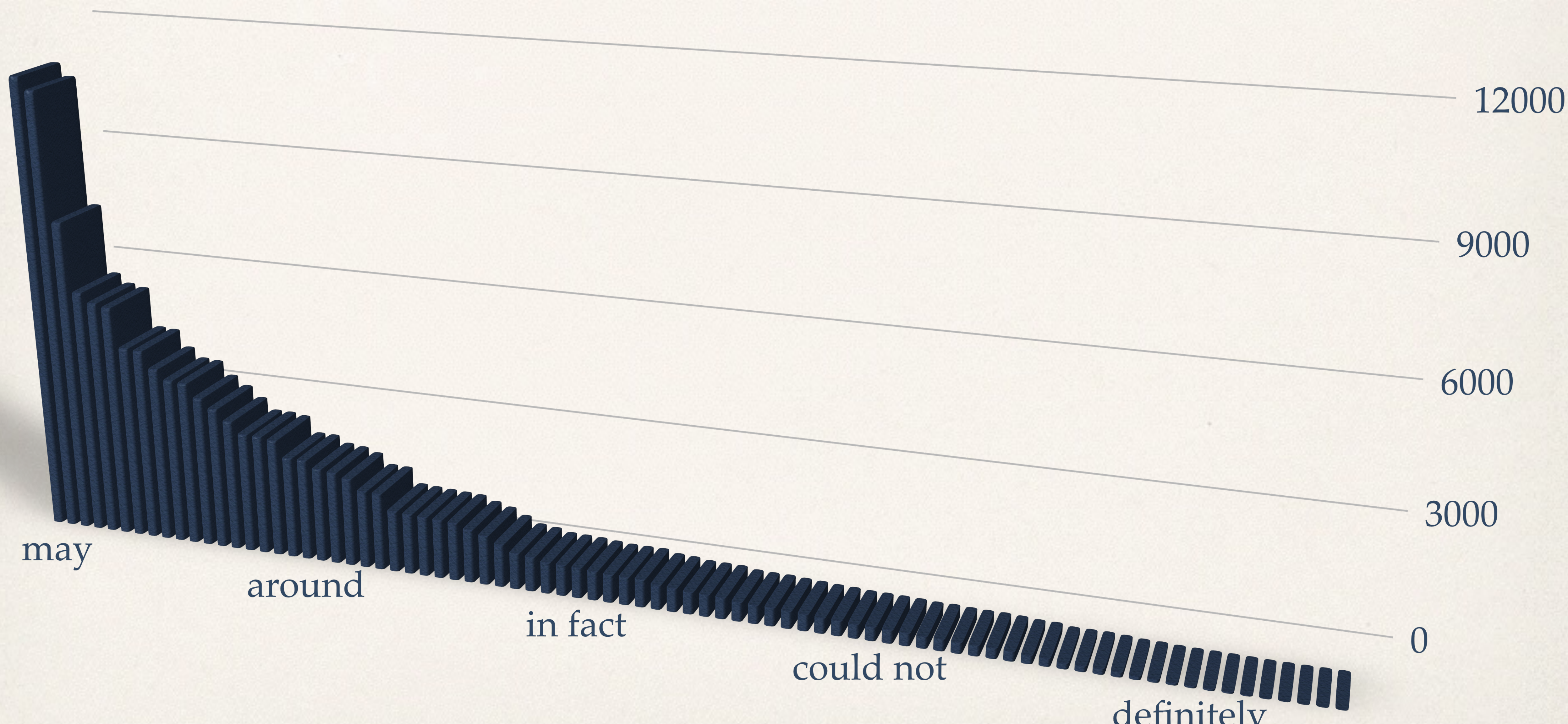
Fig. 2. Migration cell.

Epistemic language

"Modes of knowing": Communicating doubts, certainties & guesses

Modal Verbs Could Couldn't May Might Should Shouldn't Would Wouldn't Will Won't **Adjectives** Always Apparent Certain A certain extent Clear Evident Possible Probable **Nouns** Claim Doubt Estimate Evidence Possibility **Lexical Verbs** Appear Argue Assume Believe Claim Doubt Estimate Expect Indicate Know Predict Presume Propose Seem Speculate Suggest Suppose Tend Think **Adverbs** About Actually Almost Apparently Approximately Around Certainly Clearly Definitely Doubtless Essentially Evidently Frequently Generally In fact Indeed Largely Likely Never Normally Obviously Of course Often Perhaps Possibly Presumably Probably Quite Rarely Relatively Sometimes Surely Undoubtedly Usually

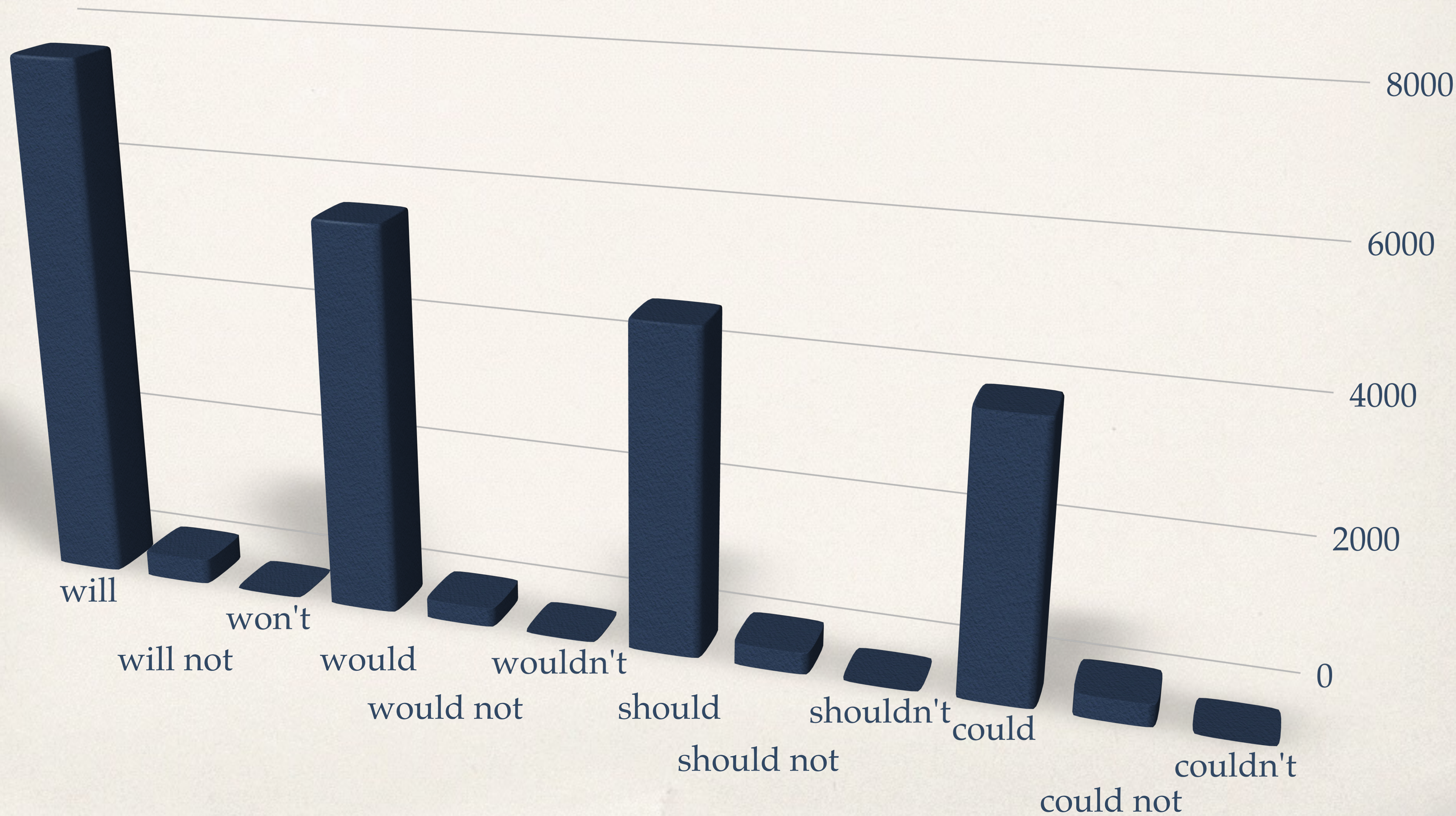
Most frequent epistemic items in academic writing (Hyland & Milton 1997)



Epistemic items in SCCERA (frequency)

may (11,127)	could (4,044)	appear (2,267)	certain (1,204)
estimate(s) (10,823)	possible (3,713)	approximately (2,160)	quite (1,041)
will (7,703)	expect (3,488)	evidence (1,902)	argue (875)
about (6,019)	predict (3,223)	might (1,889)	indeed (833)
indicate(s) (5,794)	estimate (N) (2,942)	tend (1,502)	apparent (757)
would (5,722)	likely (2,934)	clear (1,478)	wouldn't (9)
should (4,754)	relatively (2,884)	seem (1,456)	won't (6)
assume (4,727)	often (2,486)	usually (1,439)	couldn't (3)
suggest (4,315)	around (2,480)	almost (1,419)	doubtless (3)
propose (4,074)	generally (2,311)	clearly (1,302)	shouldn't (2)

Modal expressions



POS - CLAWS tagset

Position	POS tag	Info.	SCCERA	Medical	Brown
1	N	nouns	32.2%	29.1%	23.1%
2	V	verbs	13.4%	11.1%	15.5%
3	I	prepositions	13.4%	—	—
4	J	adjectives	10.2%	9.7%	6.9%

Coastal Engineering: Keywords vs. SCCERA (position)

wave (1)	storm (13)	numerical (23)	reef (33)
sea (2)	shoreline (14)	height (24)	waters (34)
coastal (3)	coast (15)	bed (25)	salinity (35)
ice (4)	erosion (16)	islands (26)	breakwater (37)
waves (5)	tidal (17)	water (27)	surge (38)
ocean (6)	tide (18)	shore (28)	swash (39)
breaking (7)	beaches (19)	offshore (29)	Atlantic (40)
beach (9)	currents (20)	island (30)	coasts (41)
shelf (10)	depth (21)	dune (31)	figure (42)
wind (12)	arctic (22)	runup (32)	shelves (44)

Coastal Engineering: Keywords vs. BNC (position)

wave (2)	ocean (14)	breaking (25)	tsunami
et al	water (16)	wind (26)	eq
coastal (5)	figure (17)	tidal (27)	boundary
ice (6)	velocity	flow	erosion
model	surface	beach (29)	the
fig	depth (20)	height	values
equation	numerical (21)	storm (32)	elevation
sea (11)	sediment (22)	level	measurements
waves (12)	shelf (23)	measured	salinity (45)
data	shoreline (24)	results	simulation

3-word clusters

based on the	with respect to	in the case
as well as	in this paper	there is a
the number of	one of the	the value of
in order to	in this study	the presence of
shown in fig	a function of	can be used
in terms of	the case of	the fact that
due to the	part of the	according to the
the effect of	a number of	as a result
the use of	the effects of	be used to
as shown in	the results of	the other hand

4-word clusters

in the case of	the results of the	it is important to
on the other hand	is shown in fig	it should be noted
as a function of	the size of the	in the context of
as shown in fig	are shown in fig	is assumed to be
as well as the	is based on the	the fact that the
can be used to	the end of the	should be noted that
on the basis of	at the end of	in the form of
with respect to the	the effect of the	it is possible to
in terms of the	at the same time	it can be seen
as a result of	in the united states	in this paper we

Keywords: Hard vs. soft sub-disciplines of civil engineering

Hard	& Structures)	**Soft**	structure dvlpt)
damping	load	project	risk
response	steel	construction	pavement
beam	force	management	risks
structural	displacement	projects	team
stiffness	strain	cost	research
bridge	equation	success	safety
control	damage	life	leadership
vibration	frequency	costs	performance
damper	excitation	managers	process

References

Hyland, K. (2008). 'As can be seen: Lexical bundles and disciplinary variation'. *English for Specific Purposes* 27: 4-21.

Hyland, K. & Milton, J. (1997). 'Qualification and certainty in L1 and L2 students' writing'. *Journal of Second Language Writing* 6(2): 183-205.

Kennedy, G. (1998). *An introduction to corpus linguistics*. New York: Longman.

Lee, D. & J. Swales (2006). 'A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora'. *English for Specific Purposes* 25: 56-75.

Pearson, J. (1998). *Terms in context.* Amsterdam: John Benjamins.

Rea Rizzo, C. (2010). 'Getting on with corpus compilation: From theory to practice'. *ESP World* 1(9): 1-22.

Scott, M. (2011). WordSmith Tools Version 6, Liverpool: Lexical Analysis Software.

Thank you!

ありがとうございました。

A brief history of Tokyo University Department of Civil Engineering

- **1914**: Department established with 4 laboratories (River & Coastal, Railways, Bridge Construction & Sanitary Engineering)
- **1923** (Great Kanto Earthquake): Earthquake & Geotechnical Engineering departments added
- **1995** (Great Hanshin Earthquake): Landscape Planning/Design & Construction Management departments added
- **2011** (Tohoku Earthquake): Flood simulation sub-department added

Keywords vs. BNC (position)

et al (2/3)	table (14)	soil (26)	ratio (36)
fig (4)	shear (15)	based (27)	spatial (37)
model (5)	using (16)	stress (28)	variables (38)
data (6)	wave (17)	the (29)	distribution (39)
equation (7)	figure (18)	observed (30)	strain (40)
results (8)	surface (19)	velocity (31)	method (41)
values (9)	parameters (20)	temperature (32)	parameter (42)
models (11)	water (22)	measured (33)	measurements (43)
flow (12)	analysis (23)	behavior (34)	shown (44)
concrete (13)	eq (25)	coefficient (35)	earthquake (45)

Epistemic items: Expressing doubt & certainty in academic writing

- “ epistemic comment is often seen as a principal means by which writers can use language flexibly to adopt positions, express points of view and signal allegiances.” (Hyland & Milton 1997: 183)

- “Our experience as EFL instructors […] lead us to believe that L2 writers find manipulation of degrees of probability particularly problematic.” (ibid: 183)

- “These problems persist for L2 writers at post graduate level where PhD supervisors are often required to counsel the need for appropriate degrees of qualification and confidence in expressing claims.” (ibid: 185)